Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **463,** 4019–4040 (2016) doi:10.1093/mnras/stw2193
Advance Access publication 2016 September 1

Lyα-emitting galaxies as a probe of reionization: large-scale bubble morphology and small-scale absorbers

Koki Kakiichi,[1]★ Mark Dijkstra,[1,2] Benedetta Ciardi[1] and Luca Graziani[1,3]

[1]*Max Planck Institute for Astrophysics, Karl-Schwarzschild Straße 1, D-85741 Garching, Germany*
[2]*Institute of Theoretical Astrophysics, University of Oslo, P.O box 1029, Blindern, 0315 Oslo, Norway*
[3]*INAF Osservatorio Astronomico di Roma, Via Frascati 33, I-00040, Monte Porzio Catone (RM), Italy*

Accepted 2016 August 31. Received 2016 August 29; in original form 2015 October 19

ABSTRACT

The visibility of Lyα-emitting galaxies during the Epoch of Reionization is controlled by both diffuse H I patches in large-scale bubble morphology and small-scale absorbers. To investigate their impacts on Lyα transfer, we apply a novel combination of analytic modelling and cosmological hydrodynamical, radiative transfer simulations to three reionization models: (i) the 'bubble' model, where only diffuse H I outside ionized bubbles is present; (ii) the 'web' model, where H I exists only in overdense self-shielded gas; and (iii) the hybrid 'web–bubble' model. The three models can explain the observed Lyα luminosity function equally well, but with very different H I fractions. This confirms a degeneracy between the ionization topology of the intergalactic medium (IGM) and the H I fraction inferred from Lyα surveys. We highlight the importance of the clustering of small-scale absorbers around galaxies. A combined analysis of the Lyα luminosity function and the Lyα fraction can break this degeneracy and provide constraints on the reionization history and its topology. Constraints can be improved by analysing the full $M_{\rm UV}$-dependent redshift evolution of the Lyα fraction of Lyman break galaxies. We find that the IGM-transmission probability distribution function is unimodal for bubble models and bimodal in web models. Comparing our models to observations, we infer that the neutral fraction at $z \sim 7$ is likely to be of the order of tens of per cent when interpreted with bubble or web–bubble models, with a conservative lower limit $\sim$1 per cent when interpreted with web models.

Key words: line: formation – radiative transfer – galaxies: high-redshift – intergalactic medium – cosmology: theory – dark ages, reionization, first stars.

1 INTRODUCTION

The Epoch of Reionization (EoR) and Cosmic Dawn are the least explored frontiers in observational cosmology and extragalactic astrophysics (Loeb & Furlanetto 2013). Galaxy surveys are one of the most important pillars of modern cosmology, allowing us to study high-redshift galaxy formation and the reionization process of the intergalactic medium (IGM). Surveys of high-redshift galaxies using Lyman-break drop-out technique (Lyman break galaxies, LBGs) (e.g. McLure et al. 2011; Ellis et al. 2013; Bouwens et al. 2015) and narrow-band filter targeting Lyα emission (Lyman Alpha Emitters, LAEs) (e.g. Malhotra & Rhoads 2004; Hu et al. 2010; Ouchi et al. 2010) have provided a deep sample of objects, indicating that reionization requires many faint galaxies below the sensitivity limit of the surveys (e.g. Robertson et al. 2013). Furthermore, observations of QSO spectra (e.g. Fan et al. 2006; Becker et al. 2015) and the cosmic microwave background (CMB) (Hinshaw et al. 2013; Planck Collaboration XVI et al. 2014; Planck Collaboration et al. 2015) offer hints that reionization is mostly completed at $z \gtrsim 6$.

However, beyond such indications, our present observational constraints on the EoR are still scarce, regarding both the reionization history and its topology/morphology. While 21-cm experiments with radio interferometers such as LOFAR,[1] MWA,[2] GMRT,[3] PAPER,[4] HERA,[5] SKA[6] offer the most direct probe of the physical state of the IGM during the EoR (e.g. Pritchard & Loeb 2012), the challenge in foreground removal and calibration remains. Substantial progress has been recently made by the 21-cm community,

★E-mail: kkakiichi@mpa-garching.mpg.de

[1] http://www.lofar.org
[2] http://www.mwatelescope.org
[3] http://gmrt.ncra.tifr.res.in
[4] http://astro.berkeley.edu/~dbacker/eor
[5] http://reionization.org
[6] http://www.skatelescope.org

but a detection is still missing. In principle, surveys of Lyα-emitting galaxies (Lyα surveys) offer an alternative and independent means from 21-cm experiments to probe the EoR and constrain the global H I fraction (e.g. see Dijkstra 2016 for review and references therein). Such an approach is attractive because of the present availability of data and up-coming surveys with the Hyper Suprime-Cam on Subaru[7] and with future telescopes such as JWST,[8] E-ELT,[9] TMT,[10] and GMT.[11] Furthermore, using multiple independent strategies can provide constraints on reionization which are less sensitive to systematic uncertainties of individual probes.

The challenge in using Lyα-emitting galaxies as a probe of reionization lies in correctly interpreting observations (Stark et al. 2010; Curtis-Lake et al. 2012; Ono et al. 2012; Treu et al. 2013; Caruana et al. 2014; Faisst et al. 2014; Konno et al. 2014; Tilvi et al. 2014; Cassata et al. 2015). The reduced visibility of Lyα emission from galaxies at $z > 6$ has already been used to infer the global H I fraction of the IGM (e.g. Dijkstra, Mesinger & Wyithe 2011; Jensen et al. 2013). However, a robust interpretation is still uncertain because of the complex radiative transfer of both ionizing and Lyα photons. The Lyα transfer involves a wide range of scales including (i) the interstellar medium (ISM), where dust and gas distribution and kinematics determine the escape fraction of Lyα photons as well as their spectral line profile (e.g. Verhamme, Schaerer & Maselli 2006; Hutter et al. 2014; Gronke et al. 2015); (ii) the circum-galactic medium (CGM), i.e. the direct environment of galaxies out to a few hundred kpc (e.g. Dijkstra, Lidz & Wyithe 2007; Zheng et al. 2010, 2011; Laursen, Sommer-Larsen & Razoumov 2011; Jeeson-Daniel et al. 2012); and (iii) the IGM, which – during reionization – contains diffuse neutral gas surrounding large ionized bubbles which themselves contain dense, self-shielding gas clouds (Bolton & Haehnelt 2013; Mesinger et al. 2015; Choudhury et al. 2015). In addition, galaxy formation physics (Forero-Romero et al. 2012) and the sample variance (Taylor & Lidz 2014) play a role in controlling the observed fraction of Lyα emission in LBGs. In order to obtain robust constraints on the global H I fraction, it is essential to understand the cosmological Lyα RT on all these scales.

The precise ionization structure of the IGM, i.e. the topology of reionization, is not only characterized by the size, abundance and distribution of large-scale ionized bubbles, but also by the small-scale dense H I absorbers self-shielded against the external ionizing sources. Interpretations of Lyα-emitting galaxies contain (often implicit) assumptions about the ionization structure of the IGM, mostly because of the difficulty to cover the entire dynamic range that is required to properly describe both the small-scale dense H I absorbers *and* the large-scale diffuse neutral IGM in reionization simulations. Two extreme assumptions, described in the first two bullets below, have been commonly adopted in the literature. Here we introduce the following terminology:

(i) *Bubble model*: in this model small-scale H I absorbers are neglected. Under this assumption, the global H I fraction measures the H I content of the diffuse neutral IGM outside ionized bubbles. We refer to this as the 'bubble model'.

(ii) *Web model*: here only the small-scale H I absorbers are considered. As this overdense gas largely traces the large-scale cosmic web, we refer to it as the 'web model'.

(iii) *Web–bubble model*: reality is a combination of the two extreme configurations above. We refer to cases that contain both neutral phases (diffuse and clumped) of gas as the hybrid 'web–bubble model'. One can visualize this as the more common bubble model, but with 'impurities' in the ionized bubbles in the form of small-scale neutral islands.

Most previous works interpreting the observed reduction in Lyα flux from $z > 6$ galaxies have favoured a very high value of H I fraction, as high as $\sim$50 per cent at $z \sim 7$ (e.g. McQuinn et al. 2007; Dijkstra et al. 2011; Jensen et al. 2013). These studies used large-scale reionization simulations which did not have the spatial resolution to resolve the self-shielded small-scale absorbers.

The lack of self-shielding gas inside ionized bubbles in large-scale reionization simulations is clearly problematic: Lyα forest observations indicate that in the post-reionized universe, i.e. $z < 5$, H I gas is locked up in damped Lyα systems (DLA) and Lyman-limit systems (LLS) (e.g. Wolfe, Gawiser & Prochaska 2005). Self-shielded absorbers (LLSs and DLAs) are also expected to reside inside ionized bubbles during reionization (see Bolton & Haehnelt 2013 for a first investigation of this point). The first investigations of hybrid web–bubble models have recently been reported (Choudhury et al. 2015; Mesinger et al. 2015). Interestingly, these papers still favour large values for the H I fraction, as high as $\sim$40 per cent at redshift $z = 7$.

In this paper we investigate the impact of large-scale patchy reionization and small-scale H I absorbers on the observed Lyα flux of galaxies, and its implication on the H I fraction measurements from Lyα surveys. We explore a unique combination of cosmological hydrodynamical, radiative transfer simulations and analytic models. Our analytic framework is powerful as it facilitates the interpretation of the results of our simulations, and provides us with a tool to quickly explore a large range of parameters describing the reionization and Lyα transfer processes in future work. This paper aims to provide a proof-of-concept of our simulation-calibrated constraints on the reionization history and topology from surveys of Lyα-emitting galaxies. Because our focus is to understand the impact of the IGM structure on the Lyα transfer, we adopt a simple model for the intrinsic properties of Lyα-emitting galaxies. In future work, we aim to provide a comprehensive methodology including both the variations of intrinsic galaxy properties and reionization.

The paper is organized as follows. Section 2 briefly reviews the cosmic history of the H I content in the universe, ranging from the EoR to the post-reionized universe. In Section 3 we present our analytic framework of cosmological Lyα radiative transfer. In Section 4 we describe the methodology employed to generate the reionization models (bubble, web, and web–bubble models), as well as the intrinsic and apparent mock catalogue of Lyα-emitting galaxies. Section 5 shows our results. The conclusions and discussion about implications on Lyα-emitting galaxy surveys are then presented in Section 6.

2 COSMOLOGICAL H I CONTENT

In this section, we review the redshift evolution of the H I content both during and after reionization. This can be quantified either by the mass-weighted $\langle f_{\rm H\,I}\rangle_{\rm M}$ or the volume-weighted $\langle f_{\rm H\,I}\rangle_{\rm V}$ neutral fraction. A compilation of current estimates in the literature is shown in Fig. 1.

[7] http://www.naoj.org/Projects/HSC/
[8] http://www.jwst.nasa.gov/
[9] http://www.eso.org/public/teles-instr/e-elt/
[10] http://www.tmt.org/
[11] http://www.gmto.org/

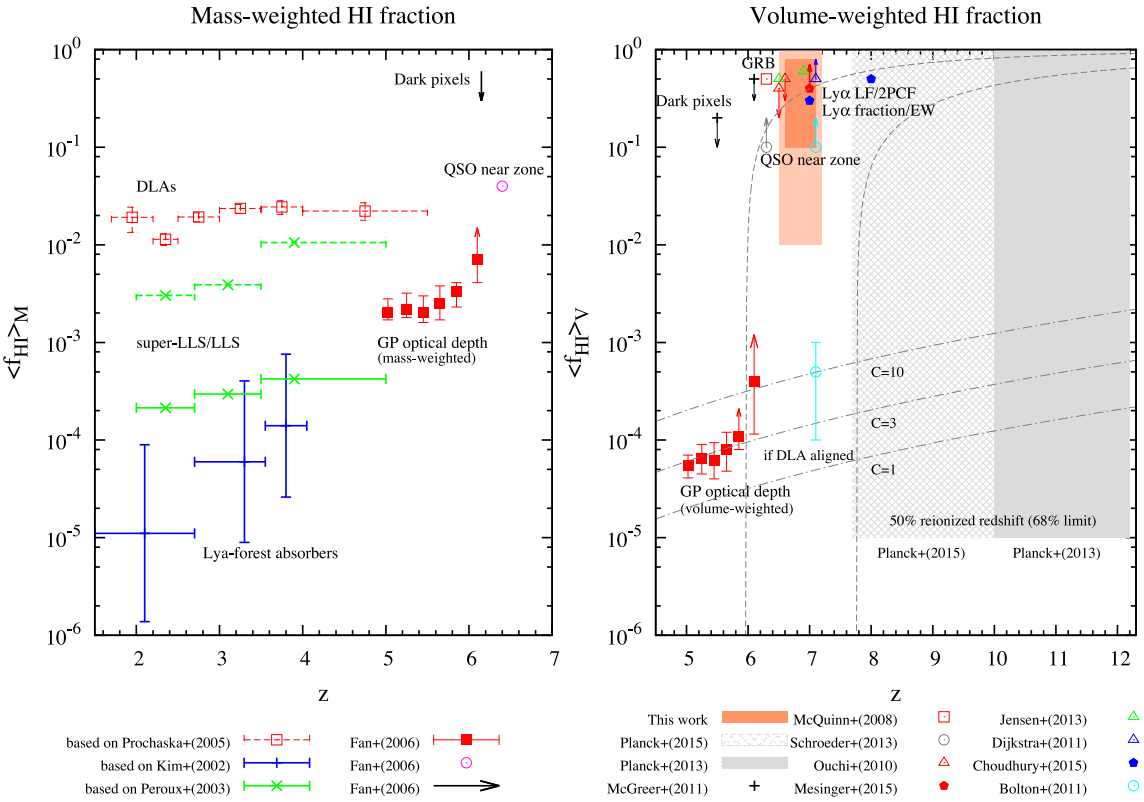

Figure 1. Cosmological H I fraction $\langle f_{\rm H\,I}\rangle_{\rm V,\,M}$ in the diffuse IGM and high-column density Lyα absorbers (LLS/DLAs) at various redshifts, from the post-reionized universe to the EoR. The mass-weighted H I fraction embedded in the small-scale absorbers is computed from equation (A2) using the fitting function to the observations of Prochaska et al. (2005), Kim et al. (2002) and Peroux et al. (2003). All the other values are compiled from the literature, as indicated by the labels above [$z = 7$ constraints by Dijkstra et al. (2011) and Jensen et al. (2013) are shifted by $+0.1$ and -0.1 in redshift to avoid a cluttering of data points). The dash–dotted lines in the right panel are the volume-weighted neutral fraction of a diffuse IGM with clumping factor $C = 1, 3, 10$ in photoionization equilibrium with a UV background $\Gamma = 10^{-12}$ s^{-1}. The dashed lines are the bubble model estimates of the neutral fraction in H I patches outside ionized bubbles (left and right lines are DEC and CONST models in Section 2.3.2).

2.1 Observational constraints on H I in the post-reionization epoch

The left panel of Fig. 1 clearly indicates that the post-reionized universe still contains neutral islands of gas in the form of self-shielding LLSs and DLAs. The abundance of the H I gas is generally quantified in terms of the H I column density distribution function (CDDF), $f(N_{\rm H\,I}, z)$, which is defined as (e.g. Wolfe, Gawiser & Prochaska 2005) $f(N_{\rm H\,I}, z) = \frac{\partial^2 \mathcal{N}}{\partial N_{\rm H\,I}\partial z}\frac{H(z)}{H_0(1+z)^2}$, where $\frac{\partial^2 \mathcal{N}}{\partial N_{\rm H\,I}\partial z}$ is the number of Lyα absorbers $\mathcal{N}(N_{\rm H\,I}, z)$ per unit H I column density $N_{\rm H\,I}$ and per unit redshift, $H(z) = H_0[\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda]^{1/2}$ and H_0 is the Hubble parameter today. The mass-weighted H I fraction embedded in small-scale absorbers is estimated from observations of $f(N_{\rm H\,I}, z)$ (see Appendix A).[12]

The left panel of Fig. 1 further shows the mass-weighted H I fraction embedded in each type of Lyα absorber. The dominant reservoir of H I gas is the high-column density Lyα absorbers, mainly DLAs. The $\langle f_{\rm H\,I}\rangle_{\rm M} \sim 1$ per cent embedded in DLAs stays approximately constant over $2 < z < 5$, while the H I fraction embedded in super-LLS and LLS, which is the second dominant H I gas reservoir, increases with redshift. The diffuse IGM, represented by the Lyα forest absorbers, is highly ionized and remains a minor reservoir of neutral gas.

2.2 Observational constraints on H I during reionization

In the right panel of Fig. 1 we have compiled various inferred values of the volume-weighed H I fraction available in the literature from CMB (Planck Collaboration et al. 2014, 2015), Gunn–Peterson optical depth (Fan et al. 2006), dark pixels (McGreer, Mesinger & Fan 2011), Gamma Ray Burst afterglow (Totani et al. 2006; McQuinn et al. 2008), quasars (QSOs) near zone (Bolton et al. 2011; Schroeder, Mesinger & Haiman 2013), Lyα luminosity function (LF; Jensen et al. 2013), equivalent width distribution (Dijkstra et al. 2011; Choudhury et al. 2015), Lyα fraction (Mesinger et al. 2015), and correlation function (Ouchi et al. 2010). We also show our suggested constraint using the Lyα LF alone (faint orange box, see Section 5.1) and when combined with the equivalent width distribution (darker orange box, see Section 5.6).

All the open points are simulation (model)-calibrated measurements, which use the Lyα radiative transfer modelling in the IGM around galaxies and QSOs. While previous works make very sensible assumptions to interpret the observed data, astrophysical systematics in such simulation (model)-calibrated measurements may raise questions about the robustness of the inferred values. While the present estimates at $5 < z < 7$ favour a volume-weighted H I fraction as high as ∼50 per cent if taken at face value, it should be

[12] Converting $f(N_{\rm H\,I}, z)$ into a constraint on the volume-weighted H I fraction, $\langle f_{\rm H\,I}\rangle_{\rm V}$, requires assumptions on the volume of LLSs and DLAs, which are model-dependent. An example of such model, and the inferred value of $\langle f_{\rm H\,I}\rangle_{\rm V}$, is discussed in Section 2.3.1.

kept in mind that these estimates are implicitly assuming a bubble model.

Interestingly, recent constraints (Choudhury et al. 2015; Mesinger et al. 2015) including both large-scale patchy reionization and small-scale absorbers still favour values for the H I fraction $\gtrsim$40 per cent at $z > 7$. Our work also prefers numbers in this range. It is the aim of this paper to understand the reason for this.

2.3 Theoretical expectations for H I

The goal of this section is to highlight the need for a hybrid web–bubble model to interpret high-z galaxy observations. We present theoretical estimates of $\langle f_{\rm H\,I}\rangle_{\rm V}$ using analytic models for the three different classes of ionization structure in the IGM. These calculations illustrate the redshift evolution of $\langle f_{\rm H\,I}\rangle_{\rm V}$ in the web (Section 2.3.1), bubble (Section 2.3.2), and hybrid web–bubble (Section 2.3.3) model.

2.3.1 H I *fraction in the web model*

In the web model, $\langle f_{\rm H\,I}\rangle_{\rm V}$ is expected to increase with increasing redshift due to decreasing photoionization rate, and/or increasing mean gas density (by Hubble expansion). $\langle f_{\rm H\,I}\rangle_{\rm V}$ can be estimated as (e.g. Miralda-Escudé, Haehnelt & Rees 2000; Bolton & Haehnelt 2007)

$$\langle f_{\rm H\,I}\rangle_{\rm V} = \int_0^{\Delta_{\rm ss}} x_{\rm H\,I}(\Delta_{\rm b})P(\Delta_{\rm b})d\Delta_{\rm b} + \int_{\Delta_{\rm ss}}^{\infty} P(\Delta_{\rm b})d\Delta_{\rm b}, \quad (1)$$

where $\Delta_{\rm b}$ is the baryon overdensity, $P(\Delta_{\rm b})$ is the volume-weighted overdensity probability distribution function,[13] and $\Delta_{\rm ss} \propto \Gamma^{2/3}$ is the density threshold above which the gas self-shields against the UV background (Schaye 2001; Furlanetto & Oh 2005). $x_{\rm H\,I}(\Delta_{\rm b}) = \alpha_{\rm A}(T)\bar{n}_{\rm H}^{\rm com}(1+z)^3 f_{\rm e}\Delta_{\rm b}/\Gamma$ is the neutral fraction obtained assuming local photoionization equilibrium with a uniform photoionization rate Γ (s^{-1}), $\bar{n}_{\rm H}^{\rm com}$ is the average comoving hydrogen number density, $\alpha_{\rm A}$ is the case A recombination rate at temperature T, and $f_{\rm e}$ is the electron fraction per hydrogen atom. The first and second term on the right hand side of equation (1) are the volume-weighted H I fraction embedded in residual H I in the diffuse IGM and the self-shielded gas, respectively.

We consider two models for the redshift evolution of the photoionization rate Γ: the CONST model assumes a constant $\Gamma = \Gamma(z = 4.75)$, while the DEC model assumes a photoionization rate decreasing with increasing redshift, i.e. $\Gamma(z) = \Gamma(z = 4.75)[(1+z)/5.75]^{-1.5}$ (Calverley et al. 2011).[14]

The blue and cyan lines in Fig. 2 show an example of the redshift evolution of $\langle f_{\rm H\,I}\rangle_{\rm V}$, with the two different contributions from residual H I in the diffuse IGM (residual H I; cyan) and neutral self-shielded gas (self-shielded H I; blue). The global H I fraction is clearly dominated at all redshifts by the self-shielded gas. While, as expected, the H I fraction increases with redshift due to the larger mean gas density, $\langle f_{\rm H\,I}\rangle_{\rm V}$ increases more markedly in the DEC model due to the lower photoionization rate.

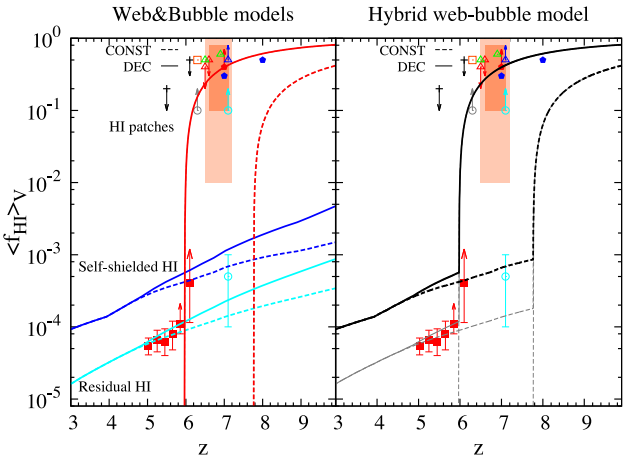

Figure 2. Redshift evolution of the volume-weighted H I fraction obtained with the analytic models. The blue and cyan lines refer to the residual neutral gas in the ionized regions (residual H I, cyan) and self-shielded gas (self-shielded H I, blue), i.e. the two different reservoirs in the web model. The red lines refer to the bubble model, where patchy neutral gas outside ionized bubbles (H I patches) is the main reservoir of H I. Results for the web–bubble model are shown in the right panel for a case in which the residual H I inside ionized bubbles is calculated including only the diffuse (grey lines) or the self-shielded (black lines) gas. CONST and DEC models are indicated as dashed and solid lines. The points in Fig. 1, inferred from observations, are overlaid as a guide, while the faint and dark orange boxes refer to our estimate based on the Lyα LF alone and used in combination with the equivalent width distribution, respectively. This plot illustrates how hybrid models are required to explain the observational constraints at all redshifts.

2.3.2 H I *fraction in the bubble model*

In the bubble model, the time evolution of the volume filling factor of ionized bubbles, the 'porosity' factor Q_i, is given by (e.g. Madau, Haardt & Rees 1999)

$$\frac{{\rm d}Q_i(t)}{{\rm d}t} = \frac{\dot{n}_{\rm ion}(t)}{\bar{n}_{\rm H}^{\rm com}} - \frac{Q_i(t)}{\bar{t}_{\rm rec,B}(t)}, \quad (2)$$

where $\dot{n}_{\rm ion}$ is the ionizing photon emissivity (in units of phs^{-1} cm^{-3}) and $\bar{t}_{\rm rec,B} = [\alpha_{\rm B}\bar{n}_{\rm H}^{\rm com}(1+z)^3 C]^{-1}$ is the case B recombination time-scale with clumping factor C. If 'empty' bubbles are assumed, we have

$$\langle f_{\rm H\,I}\rangle_{\rm V} = 1 - Q_i. \quad (3)$$

As with the web model, we consider two cases: the CONST model assumes a constant ionizing photon emissivity[15] $\dot{n}_{\rm ion}(z) = \dot{n}_{\rm ion}(z = 4.75)$, while the DEC model assumes that the ionizing photon emissivity decreases with increasing redshift, i.e. $\dot{n}_{\rm ion}(z) = \dot{n}_{\rm ion}(z = 4.75)[(1+z)/5.75]^{-1.5}$. This choice of redshift evolution is made to bracket the possible range of parameters satisfying the Fan et al. (2006) constraints.

In the left panel of Fig. 2 the redshift evolution of $\langle f_{\rm H\,I}\rangle_{\rm V}$ in the bubble model (red lines) shows a rapid change at $z \sim$ 6–8, when $\langle f_{\rm H\,I}\rangle_{\rm V}$ plummets to zero once reionization ends. A smooth transition to the post-reionized IGM, where small-scale absorbers must exist, is clearly absent from these models as no H I gas is present inside

[13] $P(\Delta_{\rm b})$ is adopted from Miralda-Escudé et al. (2000) with the redshift extrapolation of Barkana & Loeb (2004).

[14] $\Gamma(z = 4.75)$ is chosen to be consistent with the UV background measurement from the Lyα forest, i.e. $\log_{10}\Gamma(z = 4.75)/(10^{-12}\,{\rm s}^{-1}) = -0.029^{+0.156}_{-0.147}$ (Becker & Bolton 2013).

[15] $\dot{n}_{\rm ion}(z)$ is anchored at $z = 4.75$ based on Lyα forest constraints, indicating $-0.46 \lesssim \log_{10}\dot{n}_{\rm ion}/10^{51}$ ph s^{-1} cMpc^{-3} $\lesssim 0.35$ over $2.40 \lesssim z \lesssim 4.75$ (Becker & Bolton 2013). Here we assume $\dot{n}_{\rm ion}(z = 4.75) = 10^{51}$ ph s^{-1} cMpc^{-3}.

ionized bubbles. The behaviour in the CONST and DEC cases is very similar, with an earlier reionization in the former case, where a larger photoionization rate is present.

2.3.3 H I *fraction in the web–bubble model*

In the web–bubble model we assume that (i) gas inside ionized bubbles behaves as in the web model, and (ii) gas outside ionized bubbles is fully neutral. These assumptions lead to

$$\langle f_{\rm H\,I}\rangle_{\rm V} = 1 - \left[1 - \int_0^{\Delta_{\rm ss}} x_{\rm H\,I}(\Delta_{\rm b})P(\Delta_{\rm b})\mathrm{d}\Delta_{\rm b} - \int_{\Delta_{\rm ss}}^{\infty} P(\Delta_{\rm b})\mathrm{d}\Delta_{\rm b}\right] Q_i, \quad (4)$$

where the terms in square brackets are the H II fraction inside the ionized bubbles.

The redshift evolution of $\langle f_{\rm H\,I}\rangle_{\rm V}$ in the web–bubble model is shown in the right panel of Fig. 2 for a case in which the residual H I inside ionized bubbles is calculated including only the diffuse (grey lines) or the self-shielded (black lines) gas. The assumed values of the photoionization rate and ionizing photon emissivity are the same as used in the previous sections. A comparison between the left and right panels of the figure shows that the web–bubble model produces a smooth transition from the bubble model (patchy reionization) during the EoR to the web model (dominated by small-scale absorbers) in post-reionization.

Hence, to coherently explain and interpret present observations, a unified framework that includes both large-scale bubbles *and* small-scale absorbers is essential because (i) the presence of small-scale absorbers at lower-z is evident from observations (Fig. 1), and (ii) a smooth transition from a patchy reionization to a post-reionized IGM with small-scale absorbers is only possible within a hybrid web–bubble model (Fig. 2).

3 COSMOLOGICAL Lyα RADIATIVE TRANSFER

In this section we present the formalism adopted to follow the cosmological Lyα transfer through the reionization models discussed above.

The general equation describing line transfer in the Lagrangian fluid frame is (Mihalas & Mihalas 1984; Castor 2007; Meiksin 2009; Dijkstra 2014)

$$\frac{1}{c}\frac{\partial I_\nu}{\partial t} + \boldsymbol{n}\cdot\nabla I_\nu - \frac{H + \boldsymbol{n}\cdot\nabla\boldsymbol{v}\cdot\boldsymbol{n}}{c}\nu\frac{\partial I_\nu}{\partial\nu} + 3\frac{H}{c}I_\nu = -\sigma_\alpha n_{\rm H\,I}\varphi_\nu I_\nu + \sigma_\alpha n_{\rm H\,I}\int\mathcal{R}(\nu,\nu')J_\nu(\nu')\,\mathrm{d}\nu' + \varepsilon_\nu, \quad (5)$$

where I_ν is the specific intensity, J_ν is the angle-averaged intensity, ε_ν is the Lyα emissivity, $\boldsymbol{v}$ is the peculiar velocity, $\boldsymbol{n}$ is the unit direction vector of rays, $\sigma_\alpha = 0.011$ cm^{2}Hz is the Lyα cross-section, and φ_ν is the line profile of the Lyα resonance line (units Hz^{-1}). The $\boldsymbol{n}\cdot\nabla\boldsymbol{v}\cdot\boldsymbol{n}$ term is the Doppler shift effect and $\mathcal{R}(\nu,\nu')$ is the redistribution function describing the resonant scattering of Lyα photons.

There are generally no analytic solutions to equation (5). However, by performing a separation of scales, the problem can be simplified: multiple scattering effects are predominant on ISM scales because the surface brightness of Lyα photons that are scattered back into the line-of-sight at IGM scales is typically negligibly small. As scatterings on such small ISM scales can be effectively treated as a modification of the intrinsic line profile, and the scattering term can be overall neglected (Laursen et al. 2011).

Equation (5) can then be readily integrated along a line-of-sight to give the so-called '$e^{-\tau}$ approximation' (e.g. Haardt & Madau 1996; McQuinn et al. 2007; Meiksin 2009). In this approximation, the Lyα flux F_α observed from a Lyα-emitting galaxy at redshift z_s is given by

$$F_\alpha = \frac{L_\alpha}{4\pi D_{\rm L}^2(z_s)}\int S_\nu(\nu_{\rm e})e^{-\tau_\alpha(\nu_{\rm e})}\mathrm{d}\nu_{\rm e} = \frac{L_\alpha\mathcal{T}_{\rm IGM}}{4\pi D_{\rm L}^2(z_s)}, \quad (6)$$

where $\nu_{\rm e}$ is the frequency of the Lyα photon when it is emitted, $D_{\rm L}(z_s)$ is the luminosity distance, L_α is the intrinsic bolometric Lyα luminosity (in units of erg s^{-1}), $S_\nu(\nu_{\rm e})$ (in units of Hz^{-1}) is the effective intrinsic line profile (including the effect of the ISM/CGM) normalized such that $\int S_\nu(\nu_{\rm e})\,\mathrm{d}\nu_{\rm e} = 1$. $\mathcal{T}_{\rm IGM} = \int S_\nu(\nu_{\rm e})e^{-\tau_\alpha(\nu_{\rm e})}\mathrm{d}\nu_{\rm e}$ denotes the IGM transmission factor (Dijkstra et al. 2011), and the Lyα optical depth $\tau_\alpha(\nu_{\rm e})$ is

$$\tau_\alpha(\nu_{\rm e}) \approx \sigma_\alpha\int_0^{l_{\rm p}}\mathrm{d}l'_{\rm p}n_{\rm H\,I}(l'_{\rm p})\varphi_\nu\left[T, \nu_{\rm e}\left(1 - \frac{v_{\rm tot}(l'_{\rm p})}{c}\right)\right], \quad (7)$$

where T is the gas temperature and $v_{\rm tot} = H(z_s)l_{\rm p} + v(l_{\rm p})$ is the sum of the Hubble flow and the peculiar velocity. It is customary to express $\nu_{\rm e}$ in terms of a velocity shift, i.e. $\Delta v/c = 1 - \nu_{\rm e}/\nu_\alpha$. In the following we will use this convention.

We would like to note here that by using the $e^{-\tau}$ approximation we ignore photons that scatter back into the line-of-sight, which would give rise to a low surface brightness 'fuzz'. Laursen et al. (2011) compared the $e^{-\tau}$ approximation to a full Monte Carlo Lyα radiative transfer approach finding that the $e^{-\tau}$ approximation provides a good description of the transfer through the IGM as long as this is assumed to start at a distance larger than 1.5 times the virial radius of the dark matter halo hosting a Lyα galaxy. We have verified that this condition is met throughout our work.

We also introduce the mean IGM transmission factor and effective optical depth to characterize the typical impact of the intergalactic environment around Lyα-emitting galaxies. The mean Lyα flux of many Lyα-emitting galaxies is $\langle F_\alpha\rangle \approx \langle L_\alpha\rangle\langle\mathcal{T}_{\rm IGM}\rangle/(4\pi D_{\rm L}^2)$, where

$$\langle\mathcal{T}_{\rm IGM}\rangle \approx \int\langle S_\nu(\nu_{\rm e})\rangle e^{-\tau_\alpha^{\rm eff}(\nu_{\rm e})}\mathrm{d}\nu_{\rm e} \quad (8)$$

is the mean IGM transmission factor and $\tau_\alpha^{\rm eff} = -\ln\langle e^{-\tau_\alpha}\rangle$ is the effective optical depth (e.g. Haardt & Madau 1996). Here we have assumed that the intrinsic line profiles of Lyα galaxies and the optical depth of the IGM are uncorrelated, i.e. that $\langle S_\nu(\nu_{\rm e})e^{-\tau_\alpha(\nu_{\rm e})}\rangle \approx \langle S_\nu(\nu_{\rm e})\rangle e^{-\tau_\alpha^{\rm eff}(\nu_{\rm e})}$.

The optical depth contribution from different intervening IGM absorbers (the diffuse neutral IGM outside ionized bubbles and the small-scale absorbers) is additive, i.e. $\tau_\alpha = \tau_{\rm bub} + \tau_{\rm web}$. The same applies to the effective optical depth, i.e. $\tau_\alpha^{\rm eff} = \tau_{\rm bub}^{\rm eff} + \tau_{\rm web}^{\rm eff}$. In the bubble [web] model of reionization we ignore $\tau_{\rm web}$ [$\tau_{\rm bub}$], while in the web–bubble model we include both. These two terms are discussed in more detail in the following sections.

3.1 Lyα opacity from large-scale H I patches

In the bubble model the Lyα optical depth is due to diffuse expanding neutral IGM outside ionized bubbles (H I patches). The Lyα optical depth of a homogeneous H I patch extending between comoving distance from a Lyα-emitting galaxy R_1 and R_2 can be written as (Mesinger & Furlanetto 2008; Dijkstra 2014)

$$\tau_{\rm patch}(\nu_{\rm e}, R_1, R_2) = \tau_{\rm GP}\int_{x(\nu_{\rm e},R_2)}^{x(\nu_{\rm e},R_1)}\phi(x)\,\mathrm{d}x, \quad (9)$$

where $\phi = \Delta\nu_{\rm D}\varphi_\nu$ is the dimensionless line profile, $\tau_{\rm GP} = c\sigma_\alpha \bar{n}_{\rm H}^{\rm com}(1+z_s)^3/(\nu_\alpha H) \approx 4.44\times10^5[(1+z_s)/7.6]^{3/2}$ is the Gunn–Peterson optical depth, and $x(\nu_{\rm e}, R) = \{\nu_{\rm e}(1 - H(z_s)R/[(1+z_s)c]) - \nu_\alpha\}/\Delta\nu_{\rm D}$. $\Delta\nu_{\rm D} = \frac{\nu_\alpha}{c}\sqrt{\frac{2k_{\rm B}T}{m_{\rm p}}}$ is the Doppler width, with $k_{\rm B}$ Boltzmann constant and $m_{\rm p}$ proton mass.

In general, the Lyα optical depth along a line-of-sight in the bubble model is given by:

$$\tau_{\rm bub}(\nu_{\rm e}) = \sum_{i=1} \tau_{\rm patch}(\nu_{\rm e}, R_{1,i}, R_{2,i}), \tag{10}$$

where $R_{1,i}$ [$R_{2,i}$] is the near [far] side of the edge of the ith H I patch. The effective optical depth through an ensemble of H I patches is

$$e^{-\tau^{\rm eff}_{\rm bub}(\nu_{\rm e})} = \int e^{-\tau_{\rm bub}(\nu_{\rm e})} P[\tau_{\rm bub}(\nu_{\rm e})]{\rm d}\tau_{\rm bub}(\nu_{\rm e}), \tag{11}$$

where $P[\tau_{\rm bub}(\nu_{\rm e})]$ denotes the probability distribution for $\tau_{\rm bub}(\nu_{\rm e})$, which must be obtained from cosmological realizations of the bubble model.

There is a simpler limiting analytic case if we assume an ensemble of single large H I patches along all lines-of-sight. In the limit of a large H I patch ($R_2 \to \infty$), the optical depth along a line-of-sight is $\tau_{\rm patch}(\nu_{\rm e}, R_1) \approx \frac{\tau_{\rm GP}\Lambda}{4\pi^2\nu_\alpha}|\frac{\nu_{\rm e}}{\nu_\alpha}[1 - \frac{H(z_s)R_1}{(1+z_s)c}] - 1|^{-1}$, where $\Lambda = 6.25\times10^8$ s^{-1} is the damping coefficient (e.g. Miralda-Escude 1998; McQuinn et al. 2008). Then, for an ensemble of large H I patches we can evaluate the effective optical depth as

$$e^{-\tau^{\rm eff}_{\rm bub}(\nu_{\rm e})} \approx \int e^{-\tau_{\rm patch}(\nu_{\rm e}, R_1)} P(R_1)\,{\rm d}R_1, \tag{12}$$

where $P(R_1){\rm d}R_1$ is the probability to find the near side of a H I patch at a distance R_1 from a Lyα-emitting galaxy (for a related definition of bubble size distribution, see Mesinger & Furlanetto 2007). We model $P(R_1)$ as a Schechter function, $P(R_1) \propto R^{\alpha_1}\exp(-R_1/R_*)$, normalized as $\int P(R_1)\,{\rm d}R_1 = 1$; α_1 and R_* are free parameters. We compare this analytic estimate of the effective optical depth to numerical calculations in Section 5.2.

3.2 Lyα opacity from small-scale absorbers

In the web model the H I gas is distributed in a collection of self-shielded absorbers. Each absorber is characterized by its H I column density, $N_{\rm H\,I}$, and its proper velocity, v_c, relative to a given Lyα-emitting galaxy. The Lyα optical depth through a single absorber is

$$\tau_{\rm abs}(\nu_{\rm e}) = \sigma_\alpha N_{\rm H\,I}\varphi_\nu\left[T_c, \nu_{\rm e}\left(1 - \frac{v_c}{c}\right)\right], \tag{13}$$

where T_c denotes the gas temperature of an absorber.

We introduce a novel analytic model of the Lyα opacity from small-scale absorbers as follows. The effective optical depth of an ensemble of H I absorbers surrounding a Lyα-emitting galaxy is (see Appendix B for a derivation)

$$\tau^{\rm eff}_{\rm web}(\nu_{\rm e}) = \int {\rm d}N_{\rm H\,I}\frac{\partial^2\mathcal{N}}{\partial N_{\rm H\,I}\partial z}\left|\frac{{\rm d}z}{{\rm d}l_{\rm p}}\right| \times \int \frac{{\rm d}v_c}{H(z_s)}\left[1 + \xi_{\rm v}(v_c, N_{\rm H\,I})\right]\left[1 - e^{-\tau_{\rm abs}(\nu_{\rm e}, v_c, N_{\rm H\,I})}\right], \tag{14}$$

where $\xi_{\rm v}(v_c, N_{\rm H\,I})$ is the galaxy–absorber correlation function in velocity space. We refer to a Gaussian streaming model (GSM) for $\xi_{\rm v}(v_c, N_{\rm H\,I})$ when

$$1 + \xi_{\rm v}(v_c, N_{\rm H\,I}) = \int \frac{aH{\rm d}r_{12}}{\sqrt{2\pi\sigma^2_{12}(r_{12})}}(1+\xi(r_{12}))\exp\left[-\frac{(v_c - aHr_{12} - \langle v_{12}(r_{12})\rangle)^2}{2\sigma^2_{12}(r_{12})}\right], \tag{15}$$

where r_{12} is the comoving separation between a galaxy and an absorber, $\xi(r_{12})$ is the real-space galaxy–absorber correlation function, $\langle v_{12}(r_{12})\rangle$ [$\sigma_{12}(r_{12})$] is the mean radial pairwise velocity [the pairwise velocity dispersion] between galaxy–absorber pairs, and $a = (1+z_s)^{-1}$ is the scale factor.

3.2.1 *The region of influence*

As the optical depth depends on v_c, it is useful to calculate the 'critical' velocity, $v_{\rm crit}$, at which the optical depth of an absorber to a Lyα photon emitted at frequency $\nu_{\rm e}$ becomes unity for a given H I column density, i.e. $\tau_{\rm abs}(v_c = v_{\rm crit}) = 1$. In fact, to the first order, the Lyα visibility is only affected by small-scale absorbers moving away from a central Lyα-emitting galaxy with $v_c < v_{\rm crit}$. We refer to the region that contains these absorbers as the 'region of influence'. For high-column density absorbers such as LLS/DLA, the above condition is met in the wing of the absorption line profile $\varphi_\nu \approx \Lambda[4\pi^2(\nu_{\rm e}(1 - v_c/c) - \nu_\alpha)^2]^{-1}$. From the Lorentz wing it follows that for an absorber with H I column density $N_{\rm H\,I}$,

$$\frac{v_{\rm crit}}{c} = 1 - \frac{\nu_\alpha}{\nu_{\rm e}}\left(1 - \sqrt{\frac{\sigma_\alpha N_{\rm H\,I}\Lambda}{4\pi^2\nu_\alpha^2}}\right). \tag{16}$$

If we set $\nu_{\rm e} = \nu_\alpha$, then $v_{\rm crit} = c\sqrt{\frac{\sigma_\alpha N_{\rm H\,I}\Lambda}{4\pi^2\nu_\alpha^2}} = 507.3(N_{\rm H\,I}/10^{20}\,{\rm cm}^{-2})^{1/2}\,{\rm km\,s}^{-1}$. For a pure Hubble flow, the critical velocity corresponds to the comoving distance

$$D_{\rm infl} = \frac{v_{\rm crit}}{H_0}\frac{1+z}{[\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda]^{1/2}}. \tag{17}$$

As a reference, $D_{\rm infl} = 3.5(N_{\rm H\,I}/10^{20}\,{\rm cm}^{-2})^{1/2}\,h^{-1}$ cMpc at $z = 7$.

Armed with the analytic framework of Lyα transfer described above to aid the understanding of our results, in the next section we perform cosmological hydrodynamical, radiative transfer simulations and derive a mock survey of Lyα-emitting galaxies with various reionization models.

4 SIMULATIONS

In this section we describe the simulations used to model the observability of high-redshift Lyα-emitting galaxies, and the mock galaxy catalogue obtained from them.

4.1 Hydrodynamical simulations of the IGM

We employ the AMR cosmological *N*-body/hydrodynamical code RAMSES (Teyssier 2002) to simulate the IGM in a box of size $25\,h^{-1}$ cMpc on a side. The cosmological parameters are $(\Omega_{\rm m}, \Omega_\Lambda, \Omega_{\rm b}, \sigma_8, n_s, h) = (0.26, 0.74, 0.044, 0.85, 0.95, 0.72)$. The total number of dark matter particles is 256^3 with mass resolution of $m_{\rm DM} = 5.5\times10^7\,h^{-1}\,{\rm M}_\odot$, and the gas is included on the 256^3 base grid ($97.7\,h^{-1}$ckpc cell size) with two levels of refinement, reaching a 1024^3 grid ($24.4\,h^{-1}$ckpc cell size) at the highest AMR refinement level. For our choice of the simulation parameters, the cosmological Jeans length is $\sim57(T/100\,{\rm K})^{1/2}\,h^{-1}$ ckpc, which

corresponds to a Jeans mass of $\sim 1.3 \times 10^7\, h^{-1} \mathrm{M}_\odot$ at $T = 100$ K. The finest spatial resolution therefore fulfils the minimum requirement to resolve self-shielded absorbers of the order of the Jeans length (Schaye 2001). The above box size and the mass resolution are chosen as a compromise (1) to resolve the small dark matter haloes hosting the faint galaxies responsible for reionization, (2) to achieve a sufficient mass resolution for small-scale absorbers, and (3) to have an acceptable computational cost for post-processing RT simulations.

The initial conditions are generated with the COSMICS package (Bertschinger 1995). The initial temperature is set to 650 K. This is higher than the value expected from the cooling and heating of the IGM after recombination as computed from RECFAST (Seager, Sasselov & Scott 1999) to take into account the fact that our calculation did not include primordial heating mechanisms such as Compton heating by the CMB. The temperature is then calculated according to an adiabatic evolution. The initial redshift of the simulation is chosen as $z_{\rm ini} = 225$, to allow sufficient non-linearity to develop at the reionization epoch $z \sim 20$–5.

The dark matter haloes are identified using the HOP algorithm (Eisenstein & Hut 1998) as implemented in the RAMSES package.

4.2 Radiative transfer simulations

We use a two-way approach to follow the radiative transfer. In the first, we generate bubble models by performing full radiative transfer simulations of ionizing UV photons by post-processing the base AMR grid of the cosmological hydrodynamical simulation (Section 4.2.1). In the second, we generate web models by post-processing the finest AMR grid without full RT, but assuming photoionization equilibrium with a uniform UV background and a self-shielding criterion (Section 4.2.2). To generate the web–bubble models we modify the bubble models by treating the regions inside the ionized bubbles as web models (Section 4.2.3). This approach enables us to access spatial scales for the self-shielding gas which are presently beyond the computational capability of the full radiative transfer simulations.

We emphasize that the purpose of these simulations is not to produce the best possible reionization model, but to explore the impact of large-scale patchy reionization features (i.e. ionized bubbles) and small-scale absorbers on the observability of Lyα-emitting galaxies and on the inference of $\langle f_{\rm H\,I}\rangle_{\rm V}$ using Lyα surveys.

4.2.1 Bubble models

We use the cosmological radiative transfer code CRASH (Ciardi et al. 2001; Maselli, Ferrara & Ciardi 2003; Maselli, Ciardi & Kanekar 2009; Graziani, Maselli & Ciardi 2013) to generate our bubble models. For the full RT computation, we post-process the density and temperature fields on the 256^3 cells of the hydrodynamical simulation. While our box size is not sufficient to include the largest ionized bubbles present during the later stages of reionization, this does not affect the goal of the paper.

The model for the ionizing sources is based on the one described in Ciardi et al. (2012): the volume averaged ionizing emissivity, $\dot{n}_{\rm ion}$ (photons s^{-1} cMpc^{-3}), at $z > 6$ is parametrized as $\dot{n}_{\rm ion}(z) = 10^{50.89}\chi(z)\frac{\alpha_{\rm b}+3}{2\alpha}(\frac{\Gamma_{-12}(z=6)}{0.19})$, where $\chi(z) = ae^{b(z-9)}[a-b+be^{a(z-9)}]^{-1}$, with $a = 14/15$ and $b = 2/3$ (see Bolton & Haehnelt 2007). The values of $\Gamma_{-12}(z=6)$ are shown in Table 1. We assume that the ionizing emissivity is produced by galaxies with a power-law spectrum of slope $\alpha_{\rm b} = \alpha = 3$, and we distribute it among all haloes proportionally to their mass.

Table 1. List of reionization models. The columns indicate, from left to right, the model and its name, the photoionization rate in terms of $\Gamma_{-12} = \Gamma/10^{-12}$ s^{-1} as assumed at $z = 6$ [7] in the bubble [web] model, and the resulting volume-weighted H I fraction, $\langle f_{\rm H\,I}\rangle_{\rm V}$, at $z = 7$.

Bubble model	$\Gamma_{-12}(z=6)$	$\langle f_{\rm H\,I}\rangle_{\rm V}(z=7)$
B1	0.380	0.365
B2	0.190	0.676
B3	0.019	0.990
Web model	$\Gamma_{-12}(z=7)$	$\langle f_{\rm H\,I}\rangle_{\rm V}(z=7)$
W1	0.1	0.00073
W2	0.01	0.012
W3	0.005	0.032
Web–bubble model		$\langle f_{\rm H\,I}\rangle_{\rm V}(z=7)$
B1+W2		0.373
B1+W3		0.387
B2+W2		0.681
B2+W3		0.688

We ran the radiative transfer simulation using 10 gas density and temperature snapshots from $z = 15$ to $z = 5$ equally spaced in redshift, including both hydrogen and helium with a number fraction 0.92 and 0.08, respectively. For each source, we emit 10^6 photon packets distributed according to the power-law spectrum with 29 frequency bins sampled from 13.6 to 200 eV.

Finally, we produce a catalogue of bubble models for different values of $\Gamma_{-12}(z=6)$. Slices through these models are shown in the top panels of Figs 3 and 4. In Fig. 3 the maps of H I number density clearly show that the global H I fraction increases as $\Gamma_{-12}(z=6)$ decreases (from left to right), as expected. More specifically, a volume-weighted H I fraction of $\langle f_{\rm H\,I}\rangle_{\rm V} = 0.365$, 0.676 and 0.990 is obtained at $z = 7$ for $\Gamma_{-12}(z=6) = 0.380$, 0.190 and 0.019, respectively. Furthermore, the figures show the characteristic feature of patchy reionization, i.e. large-scale bubbles.

Since there is a one-to-one correspondence between $\Gamma_{-12}(z=6)$ and $\langle f_{\rm H\,I}\rangle_{\rm V}(z=7)$, we will use them interchangeably to specify the model.

4.2.2 Web models

We use the prescription of Rahmati et al. (2013) to account for self-shielding gas in the web models, which consists of a fitting function matched to their full RT transfer simulation. This prescription assumes photoionization equilibrium in each cell of the simulation with a modified background (see below), i.e.

$$\alpha_{\rm A}(T) n_{\rm H\,II} n_{\rm e} = \Gamma_{\rm Rah} n_{\rm H\,I}, \tag{18}$$

where $n_{\rm e}$ is the electron number density, and $\Gamma_{\rm Rah}$ is the modified local photoionization rate. The neutral fraction in each cell is then given by $x_{\rm H\,I} = [(\gamma+2) - \sqrt{(\gamma+2)^2 - 4}]/2$, where $\gamma = \Gamma_{\rm Rah}/(\alpha_{\rm A} n_{\rm H} f_{\rm e})$. The factor $f_{\rm e} = n_{\rm e}/n_{\rm H\,II}$ is 1 for a pure hydrogen medium, while $f_{\rm e} > 1$ if helium is included. We assume that the IGM temperature is $T \sim 10^4$K due to photoionization heating. The modified local photoionization rate is given by

$$\frac{\Gamma_{\rm Rah}}{\Gamma} = 0.98\left[1+\left(\frac{n_{\rm H}}{n_{\rm SS}}\right)^{1.64}\right]^{-2.28} + 0.02\left[1+\left(\frac{n_{\rm H}}{n_{\rm SS}}\right)\right]^{-0.84}, \tag{19}$$

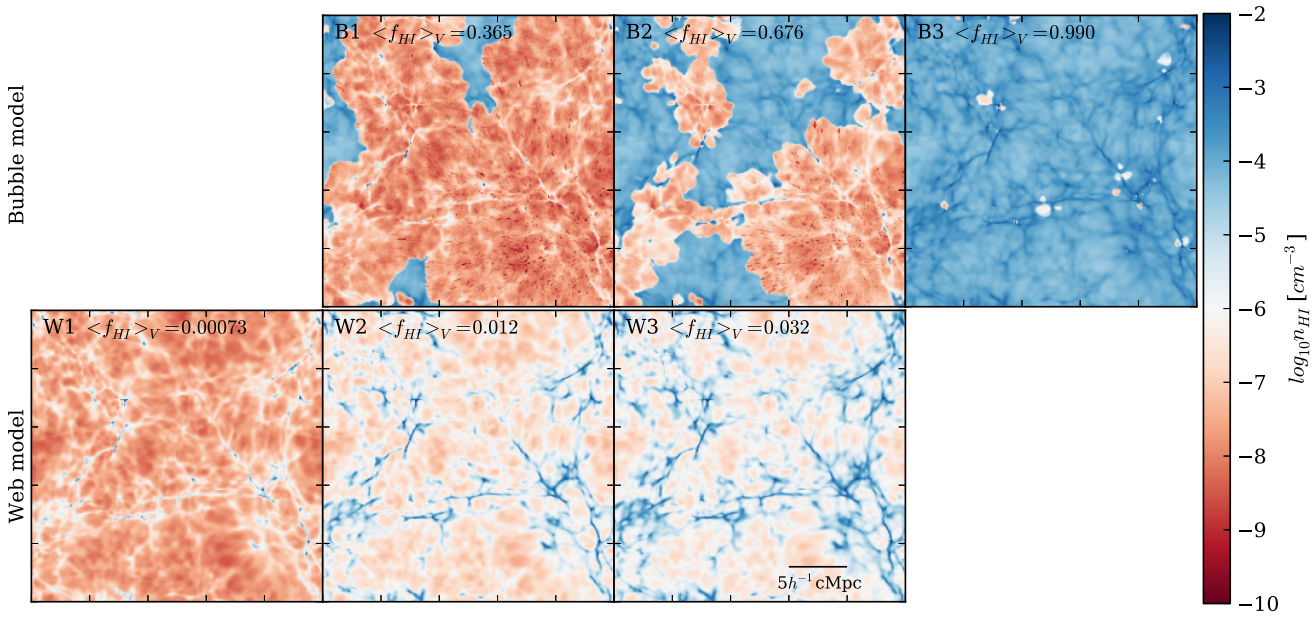

Figure 3. Neutral hydrogen number density, $n_{\rm H\,I}$, at $z = 7$ in slices of our simulations for the bubble (B1, B2 and B3; top panels) and web (W1, W2 and W3; bottom panels) models detailed in Table 1. Each snapshot is a x–y slice at 12.5 h^{-1} cMpc with 97.7 h^{-1}ckpc thickness. Panels in the same column give a similar suppression of the Lyα visibility in the observed Lyα LFs shown in Fig. 6.

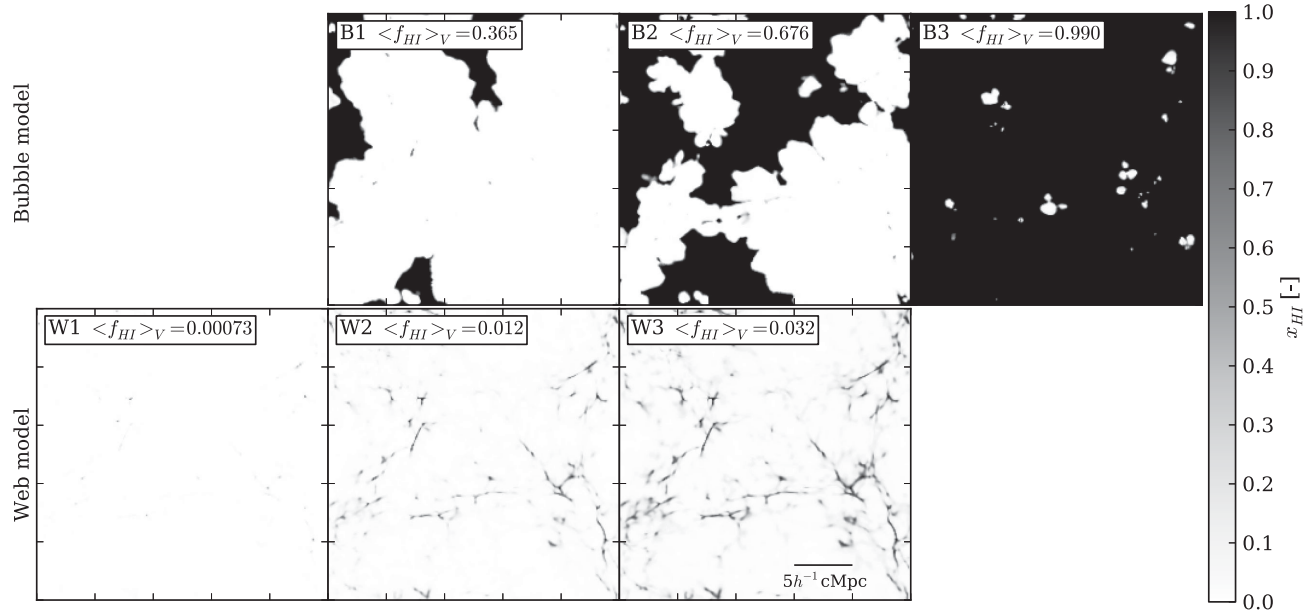

Figure 4. Same as in Fig. 3 but for the local H I fraction $x_{\rm H\,I}$.

where $n_{\rm SS}$ is the density at which the gas starts to be self-shielded

$$n_{\rm SS} = 6.73 \times 10^{-3} \Gamma_{-12}^{2/3} \left(\frac{T}{10^4\,{\rm K}} \right)^{0.17} {\rm cm}^{-3}. \quad (20)$$

To compute $n_{\rm H\,I}$ using the above prescription we use the density field of the finest AMR level 1024^3 from the hydrodynamical simulation. As a comparison, we have also calculated $n_{\rm H\,I}$ using a threshold method, in which all the cells with gas density above $n_{\rm SS}$ are assumed to be fully neutral, otherwise the neutral fraction is computed assuming photoionization equilibrium with Γ rather than $\Gamma_{\rm Rah}$, i.e. $\alpha_{\rm A}(T) n_{\rm H\,II} n_{\rm e} = \Gamma n_{\rm H\,I}$. We note that while mapping between assumed photoionization rate and the abundance of small-scale absorbers changes depending on the prescription, as long as $\langle f_{\rm H\,I} \rangle_{\rm V}$ embedded inside small-scale absorbers is similar, the result is insensitive to the self-shielding prescription. Hence, the quantity that more directly impacts the observation of Lyα-emitting galaxies is the number density of small-scale absorbers rather than the photoionization rate.

The values adopted for the uniform photoionization rate Γ_{-12} are found in Table 1. Slices through our web models are reported in the lower panels of Figs 3 and 4. Similarly to the bubble models, the maps show a higher neutral fraction for decreasing photoionization rate. However, the ionization structure of the IGM is significantly different, as the neutral gas is concentrated in high-density peaks

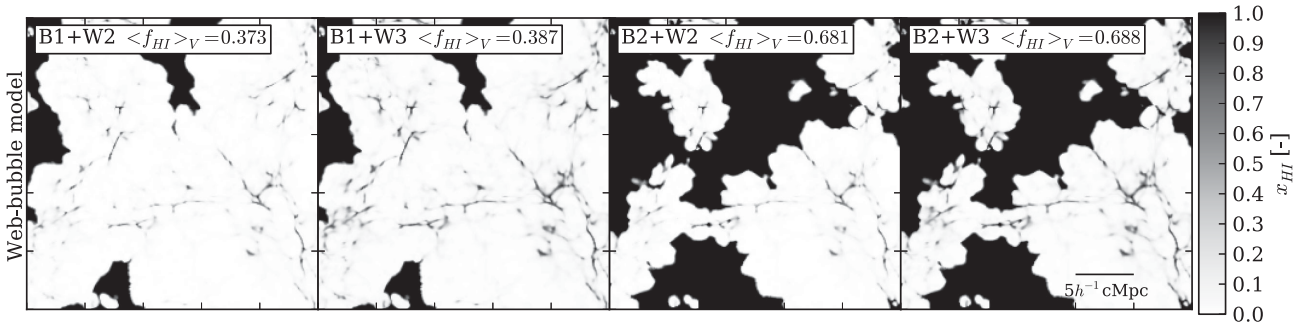

Figure 5. Slice map of the local H I fraction for hybrid web–bubble models.

where small-scale absorbers, whose distribution follows the structure of the cosmic web, reside. We note that web models are *not* equivalent to outside-in reionization scenarios. They simply show the region of the universe that is reionized early in an inside-out scenario, with residual self-shielded H I.

4.2.3 *Web–bubble models*

We generate the web–bubble models at $z = 7$ as follows. First, we take a full RT simulation used to generate the bubble models (B1 and B2). Then, we recalculate the local H I fraction inside the ionized bubbles according to the web model with a photoionization rate $\Gamma_{-12}(z = 7) = 0.01$ and 0.005 (W2 and W3) on the finest AMR grid. While in principle the photoionization rate inside bubbles is not independent of bubble size, we take this as a convenient free *ad hoc* parameter to adjust the abundance of self-shielded absorbers inside bubbles. In practice, $x_{\rm H\,I}$ is calculated locally as the maximum between the values obtained from the bubble and the web model. Our web–bubble models are catalogued in Table 1.

Slices through the web–bubble models are shown in Fig. 5 in terms of $x_{\rm H\,I}$ map. As expected, the evolution of $x_{\rm H\,I}$ with photoionization rate is the same as the one in the web and bubble models. Quantitatively, though, the neutral fraction here is slightly higher than the one in the corresponding bubble models due to the contribution of small-scale absorbers (see Table 1). In addition, the ionization structure of the IGM looks like a combination of the one from the bubble and web models, as the small-scale absorbers appear as impurities inside large-scale ionized bubbles.

4.3 Mock galaxy catalogue

The observed Lyα luminosity of a galaxy is related to its intrinsic Lyα luminosity via the IGM transmission factor $\mathcal{T}_{\rm IGM}$ as $L_\alpha^{\rm obs} = \mathcal{T}_{\rm IGM} L_\alpha$ (this is discussed more in detail in Section 3; see also Appendix D for more technical aspects). We stress that 'intrinsic' here refers to the Lyα luminosity that a galaxy would have if the IGM were transparent. As our main results are insensitive to the precise model for the intrinsic LF, we only briefly describe the methodology applied to generate the intrinsic mock galaxy catalogue.

We use the abundance matching technique (e.g. Peacock & Smith 2000; see also Appendix C) to populate dark matter haloes with Lyα-emitting galaxies. We find the relation between halo mass and intrinsic Lyα luminosity by equating the *observed* cumulative Lyα luminosity function $n(>L_\alpha)$ (in units of cMpc^{-3}) at $z = 5.7$ (Ouchi et al. 2008) to the simulated halo mass function $dn(>M'_{\rm h})/dM'_{\rm h}$ at $z = 7$,

$$n(>L_\alpha) = f_{\rm duty} \int_{M_{\rm h}(L_\alpha)}^{\infty} \frac{{\rm d}n(>M'_{\rm h})}{{\rm d}M'_{\rm h}} {\rm d}M'_{\rm h}, \quad (21)$$

where $f_{\rm duty}$ is the duty cycle and $M_{\rm h}(L_\alpha)$ is the halo mass corresponding to a Lyα luminosity L_α. We thus assume that the intrinsic Lyα LF at $z = 7$ is equal to the observed one at $z = 5.7$, and that the difference between $z = 5.7$ and $z = 7$ is entirely due to the IGM. We therefore constrain the IGM opacity using the *variation* of the Lyα LF relative to that in the post-reionization Universe. The small-scale absorbers in the post-reionized universe may affect the visibility of Lyα-emitting galaxies at $z < 6$. Hence, calibrating with $z < 6$ Lyα LF may not give a truly 'intrinsic' Lyα luminosity as defined above. This contribution should in principle be subtracted. However, as we will confirm in Section 5.3, the impact of small-scale absorbers at $z \lesssim 6$ is small. Note, though that ignoring the post-reionization optical depth of the IGM causes us to *underestimate* the intrinsic Lyα luminosity of galaxies, which in turn leads us to underestimate the IGM opacity and hence the neutral fraction in the IGM (see Dijkstra et al. 2011). The abundance matching technique gives a semi-empirical relation between the halo mass and the intrinsic Lyα luminosity for each $f_{\rm duty}$ (examples are shown in Fig. C1). In our fiducial case we use $f_{\rm duty} = 1$. We then populate each halo with a single Lyα-emitting galaxy of intrinsic Lyα luminosity given by the $M_{\rm h}$–L_α relation.

Because observations are available only down to $\log_{10}[L_\alpha/({\rm erg\,s^{-1}})] \approx 42.5$, to extend the calculations to lower luminosities we extrapolate assuming a faint-end slope of 1.5 (Ouchi et al. 2008, but see Gronke et al. 2015; Dressler et al. 2015 for both theoretical and observational support for significantly steeper slopes of ≈2.2). We note that because of the small box size (which is needed to include small-scale absorbers and the faint galaxies responsible for reionization), the simulated LFs only extend to $\log_{10}[L_\alpha/({\rm erg\,s^{-1}})] \approx 42.8$.

We model the Lyα transfer in the ISM/CGM through the Lyα spectral line profile (e.g. Dijkstra et al. 2011; Jensen et al. 2014; Choudhury et al. 2015), by assuming a Gaussian profile with a velocity dispersion $\sigma_{\rm v} = 20.4\,h^{1/3}(M_{\rm h}/10^8\,{\rm M}_\odot)^{1/3}[(1+z)/7.6]^{1/2}\,{\rm km\,s^{-1}}$ (Santos 2004; Zheng et al. 2010), shifted redward by $\Delta v = 600\,{\rm km\,s^{-1}}$ to mimic the effect of scattering through a galactic wind (Dijkstra & Wyithe 2010; Dijkstra et al. 2011). This is rather arbitrary, but Steidel et al. (2010) and Willott et al. (2015) justify a number between 200 and 800 km s^{-1}. While the quantitative results are affected by this choice, the qualitative conclusions in this paper remain valid. We point out that our model

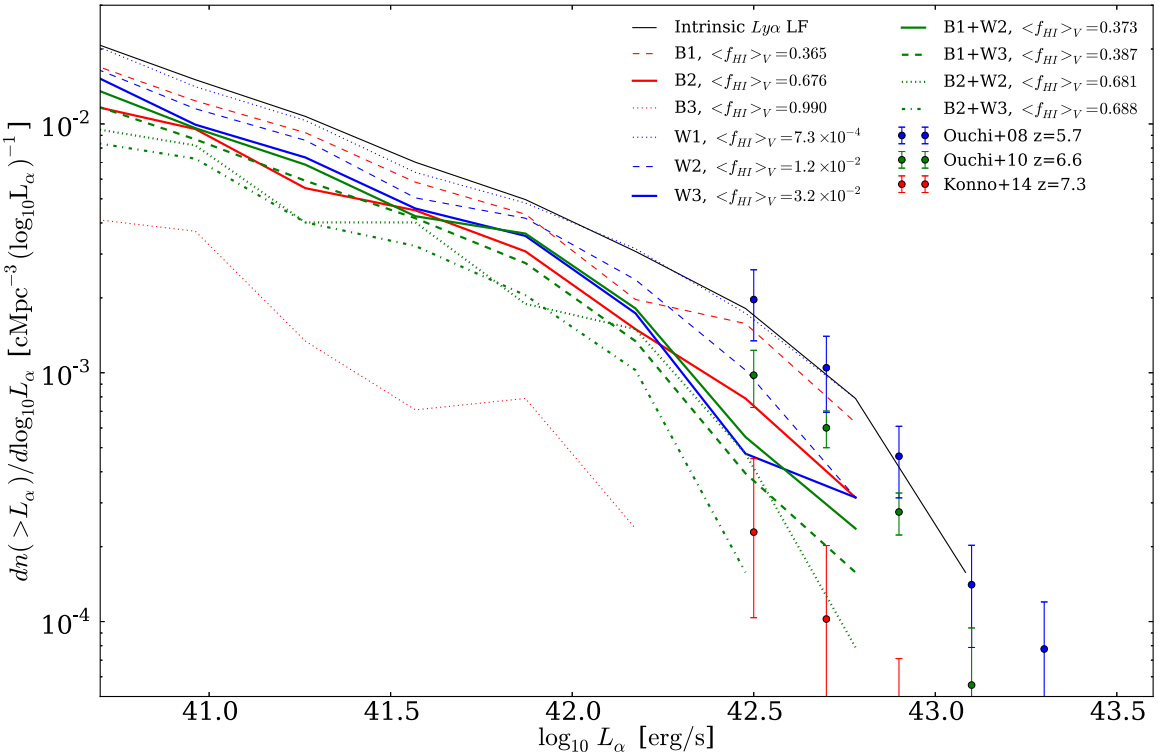

Figure 6. Intrinsic (black line) and observed differential Lyα LFs at $z = 7$ as expected for bubble (red lines), web (blue lines), and web–bubble (green lines) reionization models. The observed data points of Ouchi et al. (2008, 2010) and Konno et al. (2014) are shown as blue, green, and red points. The model of the intrinsic Lyα LF is shown as a black line. For each model, the line style refers to a different value of the parameters, as indicated in Table 1. The figure shows the degenerate impact of large-scale H I patches and small-scale absorbers on the Lyα LF.

assumes a universal line profile and shift, while a distribution is more likely. Since $\mathcal{T}_{\rm IGM}$ is highest for redshifted Lyα lines, this can bias samples of Lyα-selected galaxies to larger Δv.

5 RESULTS

5.1 Lyα luminosity function

We first show the impact of large-scale H I patches and small-scale absorbers on the Lyα LF in Fig. 6, which contains the differential intrinsic Lyα LF of galaxies (black solid line) together with the predicted apparent LF for our bubble (red lines), web (blue lines), and web–bubble (green lines) models with different values of $\langle f_{\rm H\,I}\rangle_{\rm V}$. Fig. 6 shows the following.

(i) The predicted LF decreases with $\langle f_{\rm H\,I}\rangle_{\rm V}$ as naturally expected, because more neutral hydrogen in the universe increases the overall opacity to Lyα photons.

(ii) The relative abundance of large-scale bubbles and small-scale absorbers is a key factor to estimate the observed Lyα LF. The bubble, web, and web–bubble models predict almost identical LFs for vastly different $\langle f_{\rm H\,I}\rangle_{\rm V}$. For example, a bubble model with $\langle f_{\rm H\,I}\rangle_{\rm V} = 0.676$ (B2) gives rise to a LF that is practically indistinguishable from that of a web model with $\langle f_{\rm H\,I}\rangle_{\rm V} = 0.032$ (W3) or of a web–bubble model with $\langle f_{\rm H\,I}\rangle_{\rm V} = 0.373$ (B1+W2). This was first pointed out by Bolton & Haehnelt (2013).

(iii) The presence of small-scale absorbers inside ionized bubbles provides an opacity additional to that from the neutral patches between large-scale bubbles. This is clear comparing e.g. the LFs from B1 (dashed red line) to those from B1+W2 (dotted green) or B1+W3 (dashed green).

(iv) Web models with $\langle f_{\rm H\,I}\rangle_{\rm V} \sim 10^{-2}$ correspond to bubble models with $\langle f_{\rm H\,I}\rangle_{\rm V} \sim 10^{-1}$. Table 1 indicates that this requires $\Gamma \leq 10^{-14}$ s^{-1}. For example, the red dashed (B1) and blue dashed (W2) lines in Fig. 6 show that $\Gamma_{-12}(z = 7) = 0.01$ is needed for a web model to produce a LF similar to that of a bubble model with $\langle f_{\rm H\,I}\rangle_{\rm V} \sim 0.4$. This is in agreement with Mesinger et al. (2015).

(v) Comparing the simulations to the observations of Ouchi et al. (2008, 2010) and Konno et al. (2014), we conclude that at $z = 6.6$, 40 per cent $\lesssim \langle f_{\rm H\,I}\rangle_{\rm V} \lesssim$ 70 per cent for the bubble model, $\langle f_{\rm H\,I}\rangle_{\rm V} \sim$ 1 per cent for the web model, and $\langle f_{\rm H\,I}\rangle_{\rm V} \lesssim$ 40 per cent for a web–bubble model. At $z = 7$ we have instead 70 per cent $\lesssim \langle f_{\rm H\,I}\rangle_{\rm V} \lesssim$ 99 per cent, $\langle f_{\rm H\,I}\rangle_{\rm V} \gtrsim$ 3 per cent, and 40 per cent $\lesssim \langle f_{\rm H\,I}\rangle_{\rm V} \lesssim$ 70 per cent, respectively. The inferred H I fraction thus highly depends on the reionization model adopted.

While the aim of this paper is to present a proof of concept and we defer to future work more rigorous and precise constraints on the H I fraction, these results are in excellent agreement with existing work (Bolton & Haehnelt 2013; Choudhury et al. 2015; Mesinger et al. 2015) and underline the importance of understanding the precise ionization structure of the IGM during the EoR in terms of both large-scale bubble features and small-scale absorbers. In the following, we use the simulations described in Section 4 and the analytic formalism outlined in Section 3 to gain more insight into the Lyα RT and the inference of $\langle f_{\rm H\,I}\rangle_{\rm V}$ from observed Lyα-emitting galaxies.

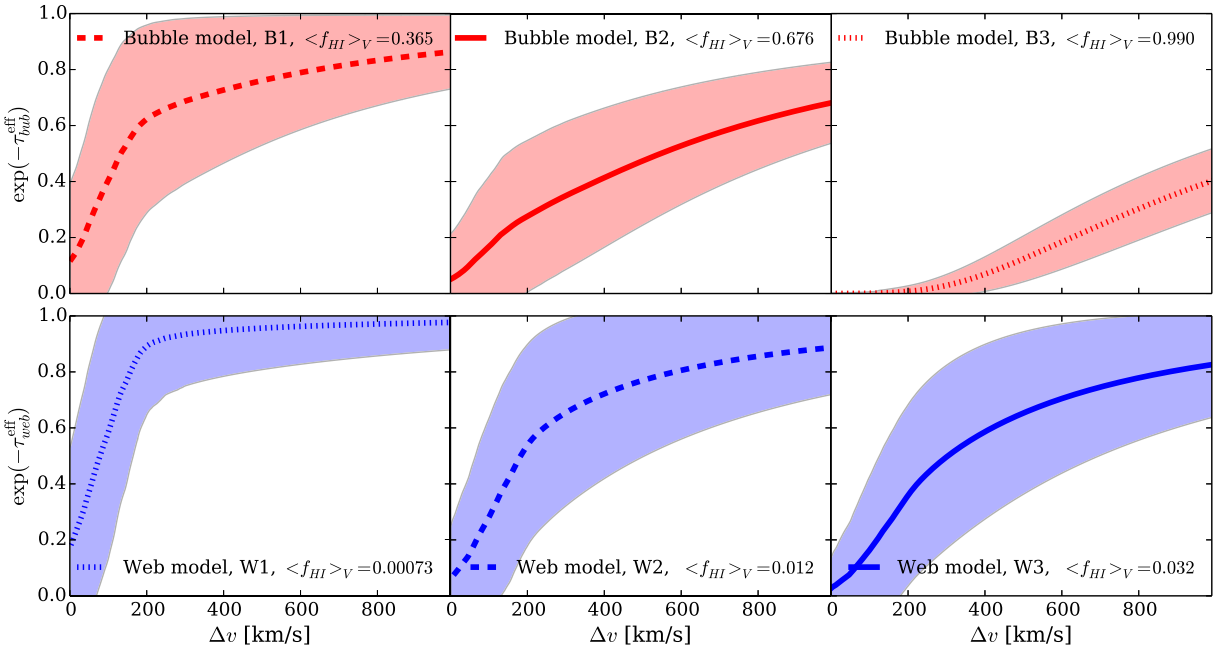

Figure 7. The mean Lyα damping wing opacity profiles in bubble and web models as a function of velocity shift relative to the line centre. A higher velocity shift corresponds to the redward (longer wavelength) of Lyα. The effective optical depths are extracted from simulations of the bubble (B1, B2 and B3; top panels) and web (W1, W2 and W3; bottom panels) models at $z = 7$. The shaded regions indicate the 1σ scatter around the mean.

5.2 The red damping wing in bubble models

We now analyse the Lyα red damping wing opacity to quantify the impact of large-scale H I patches on the visibility of Lyα-emitting galaxies. The red lines in the top panel of Fig. 7 show the mean transmission $\exp(-\tau_{\rm bub}^{\rm eff})$ as a function of Δv for three different values of $\langle f_{\rm H\,I}\rangle_{\rm V}$ (B1, B2, and B3 from left to right). We evaluate the effective optical depth directly as an average of $e^{-\tau_{\rm bub}}$ using line-of-sight skewers from galaxies extracted from the simulations. The shaded region indicates the 1σ dispersion $\sigma^2_{\tau_{\rm bub}} = \langle (e^{-\tau_{\rm bub}} - e^{-\tau^{\rm eff}_{\rm bub}})^2 \rangle$. We have used 1185 lines-of-sight, i.e. equivalent to the number of galaxies in the simulation box.

The damping wing becomes more opaque with increasing neutral fraction and decreasing Δv. The opacity varies significantly between different lines-of-sight as indicated by the large dispersion of $\sigma_{\tau_{\rm bub}} \sim 0.2$.

5.2.1 *Comparison to the analytic model*

To see how well the red damping wing can be captured by the analytic approximation, in Fig. 8 we compare the results from our B2 model to those obtained using equation (12) with $\alpha_1 = 0.5$ and $R_* = 1.7, 3.0, 5.0, 10.0\,h^{-1}$ cMpc. Note that the case with $R_* = 1.7$ (thickest black dashed line) represents the Schechter function fit to the simulated $P(R_1)$ distribution. Fig. 9 shows that the Schechter function fit is indeed a good approximation to the simulation, in which the distance to the near-side of the closest H I patch peaks at $\sim 5\,h^{-1}$ cMpc from a galaxy.

The comparison in Fig. 8 clearly indicates that the analytic model is too crude to capture the red damping wing behaviour found in the simulations, and systematically overestimates the optical depth, although the bubble size distribution is modelled reasonably well. The discrepancy highlights that the opacity is coming indeed from the neutral gas distributed among multiple H I patches, rather than in a single large H I patch, as assumed in the analytic model of equation (12). This in fact leads to an overestimate of the neutral gas and thus of the opacity.

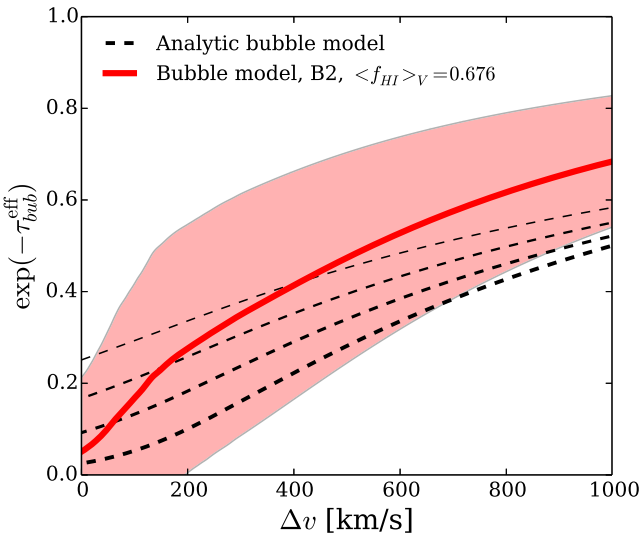

Figure 8. Comparison between the analytic and simulated bubble model. The black dashed lines are the effective optical depth from the analytic approximation (equation (12)) with a fixed slope $\alpha_1 = 0.5$ and $R_* = 1.7, 3.0, 5.0, 10.0\,h^{-1}$ cMpc (lines from bottom to top). The discrepancy between the simulation and the analytic model is due to the large single H I patch approximation used in the latter. See text for detail.

In addition, the single large H I patch approximation is also responsible for a different shape of the damping wing, because the optical depth scales as Δv^{-1} (e.g. Miralda-Escude 1998). On the other hand, the presence of multiple ionized bubbles in the simulations makes the medium more transparent, and hence the damping wing profile steeper. This implies that, unless the analytic approximation is improved to take into account the complex ionized

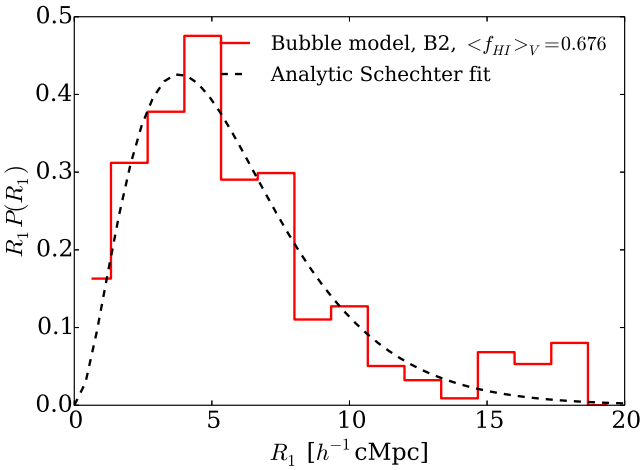

Figure 9. Distribution of the distance, R_1, to the near-side of H I patches from Lyα-emitting galaxies. The red line shows the result for the bubble model simulation B2. The black dashed line is the Schechter function fit to the simulated $P(R_1)$, where the best-fitting parameters are $\alpha_1 = 0.49$ and $R_* = 1.7\,h^{-1}$ cMpc.

bubble distribution, (semi-)numerical simulations of patchy reionization are required to properly model the Lyα opacity in the diffuse neutral IGM.[16]

5.3 The red damping wing in web models

The ensemble of small-scale absorbers can also form a damping wing feature in the effective optical depth towards Lyα galaxies as shown in the bottom panel of Fig. 7, where the blue lines and the shaded areas refer to the mean transmission $\exp(-\tau^{\rm eff}_{\rm web})$ and to the 1σ dispersion $\sigma^2_{\tau_{\rm web}} = \langle (e^{-\tau_{\rm web}} - e^{-\tau^{\rm eff}_{\rm web}})^2 \rangle$ as a function of Δv.

Similarly to the bubble model, Fig. 7 indicates that the damping wing in web models becomes more opaque with increasing neutral fraction and decreasing Δv. Neutral fractions $\langle f_{\rm H\,I} \rangle_{\rm V} \sim 10^{-2}$ (W2 and W3), i.e. much higher than the one in the post-reionized universe (which is $\langle f_{\rm H\,I} \rangle_{\rm V} \sim 10^{-4}$), are required to produce an ∼60–80 reduction of Lyα visibility at $\Delta v = 600\,{\rm km\,s^{-1}}$. On the other hand, the effective optical depth in W1 (which has a neutral fraction closer to $\sim 10^{-4}$) is $e^{-\tau_{\rm eff}} > 0.9$ at $\Delta v = 600\,{\rm km\,s^{-1}}$, i.e. it hardly affects the Lyα visibility.[17] The scatter around the effective optical depth is again large, with $\sigma_{\tau_{\rm web}} \sim 0.2$.

Finally, a comparison between the effective optical depth in the web and bubble models (e.g. B2 versus W3 in Fig. 7) shows that small-scale absorbers can produce a profile and scatter of the red damping wing similar to those of the bubble models. This explains the similarity in the Lyα LFs observed through the large-scale bubbles and small-scale absorbers.

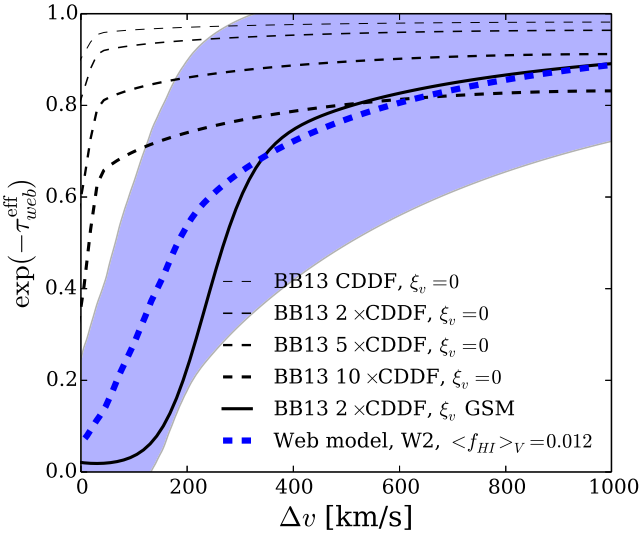

Figure 10. Comparison between the analytic and simulated web model. The black dashed lines are the effective optical depth from the analytic approximation (equation (14) using the BB13 CDDF without the velocity–space galaxy–absorber correlation function, i.e. $\xi_{\rm v} = 0$. The black solid line instead includes a GSM for $\xi_{\rm v}$. This figure demonstrates the importance of the velocity–space galaxy–absorber correlation function.

The impact of the assumed Lyα velocity offset of the line profile can be quantified from Fig. 7. At our fiducial choice of $600\,{\rm km\,s^{-1}}$ velocity offset, the effective opacity for the bubble model (B2) is $\exp(-\tau^{\rm eff}_{\rm bub}) \sim 0.53$. Similarly for the degenerate web model (W3), the effective opacity is $\exp(-\tau^{\rm eff}_{\rm web}) \sim 0.70$. These similar effective opacities produce a degenerate Lyα LF in the two models. By changing the velocity offset to e.g. $400\,{\rm km\,s^{-1}}$, the opacities of the two models increase to $\exp(-\tau^{\rm eff}_{\rm bub}) \sim 0.42$ (for bubble model) and $\exp(-\tau^{\rm eff}_{\rm web}) \sim 0.58$ (for web model). While the opacity is lowered by ∼0.1 compared to the $600\,{\rm km\,s^{-1}}$ case the relative difference between bubble and web model is about the same. This means that while the LF is further suppressed because the emergent Lyα line is closer to the line centre, the impact of the different Lyα velocity offset ($600\,{\rm km\,s^{-1}}$ versus $400\,{\rm km\,s^{-1}}$) does not modify the degeneracy.

5.3.1 *Comparison to the analytic model*

Fig. 10 compares the simulation and the analytic effective optical depth described by equation (14) in Section 3.2. The black dashed lines refer to the analytic model without the effect of clustering and velocity field, i.e. $\xi_{\rm v} = 0$ in equation (14), while the black solid line uses the GSM for $\xi_{\rm v}$, i.e. equation (15). The analytic model employs a factor of 2–10 boost to the extrapolated CDDF fit of Becker & Bolton (2013) at $z \simeq 7$ (hereafter BB13 CDDF) to mimic the rapidly increasing abundance of small-scale absorbers. Our fiducial value is 2 × CDDF (see Section 5.3.2, a discussion on the reason of this choice).

Fig. 10 shows clearly that we cannot reproduce the results from our simulation by *only* changing the CDDF amplitude, while the agreement is much better if we simultaneously change the CDDF amplitude *and* include the galaxy–absorber correlation function in velocity space (see Section 5.3.3 for the reason of the discrepancy

[16] One obvious improvement of the analytic model would be to introduce an outer radius R_2 for the first diffuse neutral patch, and construct a PDF for R_2 which can then be included into equation (12) to give

$$e^{-\tau^{\rm eff}_{\rm bub}(v_{\rm e})} \approx \int P(R_1)\,{\rm d}R_1 \int e^{-\tau_{\rm patch}(v_{\rm e},R_1,R_2)} P(R_2|R_1), \qquad (22)$$

where $P(R_2|R_1)$ denotes the conditional probability of R_2, given R_1. We have started to include such improvement in our model. However, due to the difficulty in finding an analytic fitting function for $P(R_2|R_1)$, we have deferred this to a future work.

[17] This justifies the calibration of the intrinsic model discussed in Section 4.3. However, the left bottom panel of Fig. 7 shows that $\exp(-\tau^{\rm eff}_{\rm web}) \sim 0.6$ at $\Delta v \sim 100\,{\rm km\,s^{-1}}$ even with a neutral fraction as small as the one in the post-reionized universe. For galaxies that have $\Delta v < 200\,{\rm km\,s^{-1}}$ the impact of small-scale absorbers at $z < 6$ should therefore be taken into account.

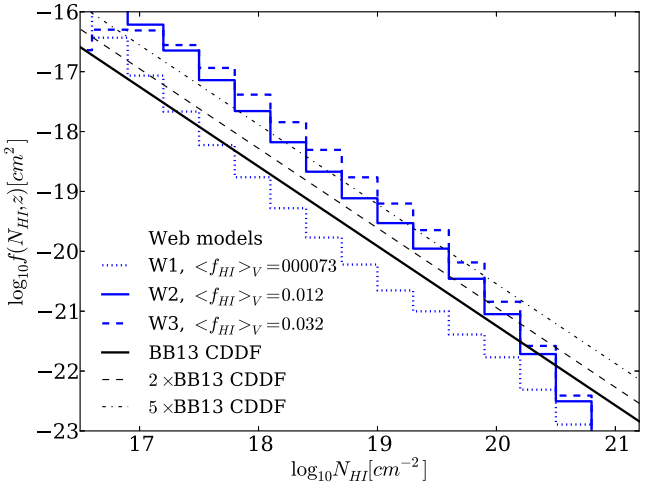

Figure 11. H I CDDFs extracted from web model simulation W1 (blue dotted line), W2 (blue solid), and W3 (blue dashed). The black lines refer to the extrapolation to $z \simeq 7$ of the Becker & Bolton (2013) fit (BB13 CDDF; solid), multiplied by a factor of 2 (dashed) and 5 (dotted).

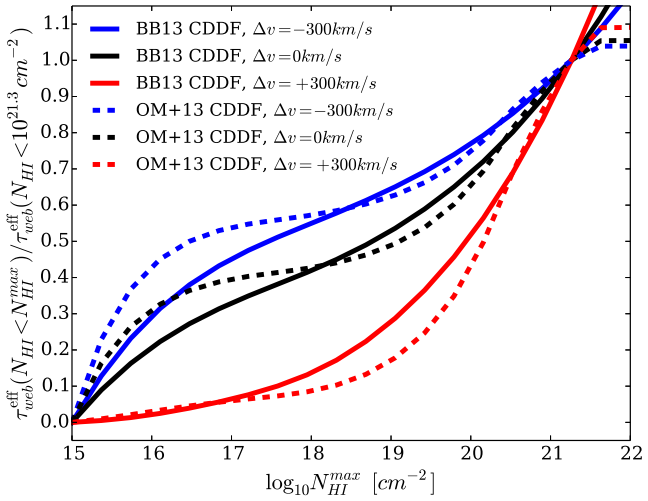

Figure 12. Cumulative contribution to $\tau^{\rm eff}_{\rm web}$ as a function of the maximum cut off column density $N^{\rm max}_{\rm H\,I}$. Three different colours correspond to the optical depth redward ($\Delta v = 300\,{\rm km\,s^{-1}}$, red line), at line centre ($\Delta v = 0\,{\rm km\,s^{-1}}$, black line), and blueward ($\Delta v = -300\,{\rm km\,s^{-1}}$, blue line). The two functional forms of H I CDDF by Becker & Bolton (2013) (solid lines) and O'Meara et al. (2013) (dashed lines) are plotted. This figure shows that the optical depth redward of line centre, i.e. red damping wing, is mostly sensitive to strong H I absorbers with $N_{\rm H\,I} > 10^{19}\,{\rm cm^{-2}}$, whereas the optical depth at smaller Δv is increasingly more affected by weaker H I absorbers with $N_{\rm H\,I} < 10^{17-19}\,{\rm cm^{-2}}$.

at $\Delta v < 200\,{\rm km\,s^{-1}}$). In other words, both the abundance of small-scale absorbers and their velocity–space clustering around galaxies play a key role in determining the Lyα visibility. In the following sections, we discuss in more detail the impact of (i) changing the CDDF (Section 5.3.2) and (ii) the galaxy–absorber clustering (Section 5.3.3).

5.3.2 *CDDF and* $N_{\rm H\,I}$*-dependence of the optical depth*

We first justify the artificial boosting factor of the power-law CDDF adopted in the analytic model. Fig. 11 compares the CDDF obtained in our web model simulations[18] to the BB13 CDDF with a factor of 1, 2 and 5 boost. The adopted boosts broadly mimic the increase in simulated CDDF amplitude due to lower photoionization rate/higher neutral fraction ($\Gamma = 10^{-14}$ s^{-1} and 5×10^{-15} s^{-1} for W2 and W3), although the slope is not properly reproduced. The fiducial choice of 2 (corresponding to $\Gamma \sim 10^{-14}\,{\rm s^{-1}}$) approximately represents the CDDF amplitude in the range $10^{19}\,{\rm cm^{-2}} < N_{\rm H\,I} < 10^{20.7}\,{\rm cm^{-2}}$, which gives the highest contribution to the red damping opacity.

This predominance can be clearly seen in Fig. 12, which shows the ratio between the analytic effective optical depth from absorbers with column density below $N^{\rm max}_{\rm H\,I}$ and below $N^{\rm max}_{\rm H\,I} = 10^{21.3}\,{\rm cm^{-2}}$, in a case with[19] $\xi_{\rm v} = 0$. More than 80 per cent of the optical depth redwards of line centre ($\Delta v = 300\,{\rm km\,s^{-1}}$) comes from absorbers with $N_{\rm H\,I} > 10^{19}\,{\rm cm^{-2}}$, because of their prominent damping wings. On the other hand, at $\Delta v = -300\,{\rm km\,s^{-1}}$ (i.e. blueward of the line resonance) lower column density absorbers with $N_{\rm H\,I} < 10^{18}\,{\rm cm^{-2}}$ can contribute ∼50 per cent to $\tau^{\rm eff}_{\rm web}$ via resonant absorption.

This strong dependence of the optical depth on the column density of absorbers is insensitive to the assumption about the shape of the CDDF, as shown by a comparison between the solid and dashed lines in Fig. 12, which refer to models using a CDDF from Becker & Bolton (2013) and O'Meara et al. (2013), respectively. In both cases $N_{\rm H\,I} \gtrsim 10^{19}\,{\rm cm^{-2}}$ absorbers dominate the red damping wing, with a difference of only ∼10 per cent.

Hence, the red damping wing opacity mainly depends on the abundance of strong H I absorbers, e.g. super-LLSs and DLAs, around Lyα-emitting galaxies. Their rapid increase (stronger than that expected from a simple extrapolation to $z > 6$ of lower-z CDDF) provides a large red damping wing opacity.

5.3.3 *Galaxy–absorber correlation function in velocity space*

The galaxy–absorber correlation function in velocity space, $\xi_{\rm v}$, is another key factor in the formation of the red damping wing. In the GSM of equation (15), $\xi_{\rm v}$ depends on both (i) the real-space correlation function, $\xi(r_{12})$, and (ii) the galaxy–absorber pairwise mean velocity field $\langle v_{12}(r_{12})\rangle$, and pairwise velocity dispersion $\sigma_{12}(r_{12})$.

The simulated real-space galaxy–absorber correlation function at $z = 7$ is shown in Fig. 13, together with the LBG–DLA correlation function observed by Cooke et al. (2006a,b) at $z \sim 3$. The simulated $\xi(r_{12})$ is obtained by correlating the position of the galaxies and of the cells with $x_{\rm H\,I} > 0.9$ (which represent for us self-shielded absorbers) using the $\xi(r_{12}) = DD/RR - 1$ estimator (Davis & Peebles 1983). Clustering of self-shielding gas in the vicinity of Lyα-emitting galaxies is clearly important, and the simulated real-space correlation function appears (maybe surprisingly) similar to its lower-redshift observed counterpart. We thus adopt the Cooke et al. (2006b) correlation function for our GSM in Fig. 10.

The mean pairwise velocity between Lyα-emitting galaxies and absorbers defined above is shown in the top panel of Fig. 14, in terms of both the proper peculiar velocity $\langle v_{12}(r_{12})\rangle$ (blue lines) and of the total proper velocity $H(z_s)r_{12}/(1+z_s) + \langle v_{12}(r_{12})\rangle$ (cyan lines). The solid black line is the best-fitting curve to the mean pairwise velocity,

[18] We have computed the CDDF by taking the projected column density over 10 cells. The highest $N_{\rm H\,I}$ bins ($\log_{10}N_{\rm H\,I}/{\rm cm^{-2}} \sim 21.3$ and 20.6) are about ∼0.3 dex larger than those calculated with a single cell, but the numbers converge for larger projected lengths. The effect is minor in the other bins.

[19] The inclusion of the velocity–space correlation function, for example inflowing low column density absorbers, would enhance the contribution of lower column density absorbers to the optical depth.

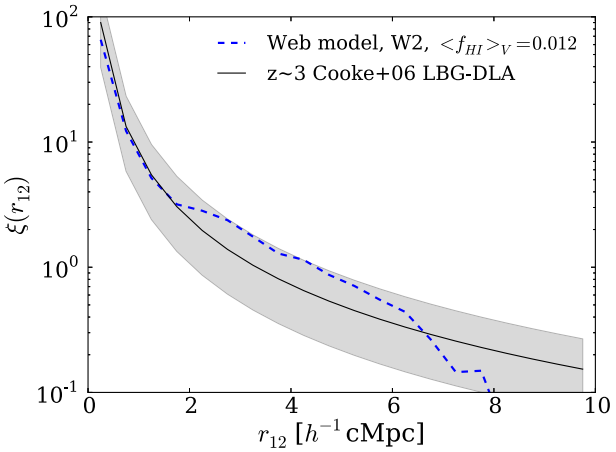

Figure 13. Simulated real-space galaxy–absorber correlation function at $z = 7$ (dashed blue line) and best fit to the LBG–DLA cross-correlation function observed by Cooke et al. (2006b) at $z \sim 3$ (solid black). The latter is $\xi(r_{12}) = (r_{12}/r_0)^{-\gamma}$, where $r_0 = 3.32 \pm 1.25\,h^{-1}$ cMpc, and $\gamma = 1.74 \pm 0.36$. The shaded region spans the upper and lower 1σ errors in the observed correlation length for a fixed slope $\gamma = 1.74$. The figure demonstrates the clustering of small-scale absorbers around galaxies.

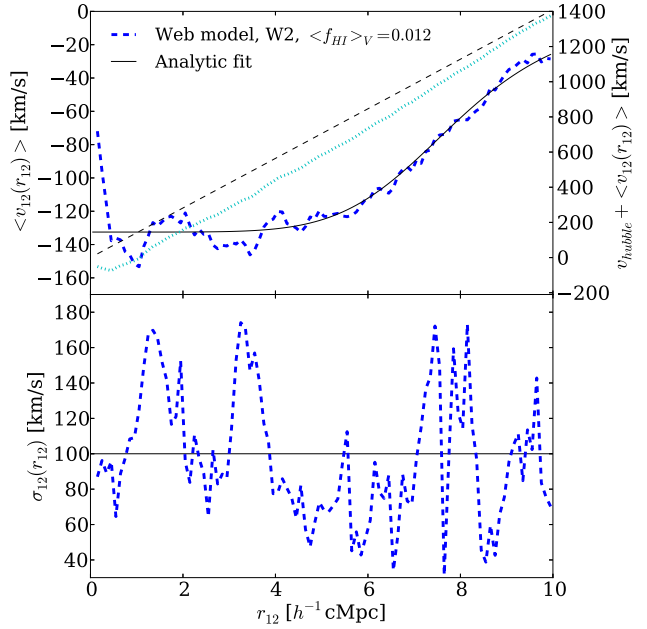

Figure 14. Galaxy–absorber mean pairwise velocity (top panel) and pairwise velocity dispersion (bottom panel) at redshift $z = 7$. The blue line (with left y-axis) shows the proper mean pairwise peculiar velocity $\langle v_{12}(r_{12})\rangle$ and pairwise velocity dispersion $\sigma_{12}(r_{12})$ between absorbers and galaxies in simulation W2, while the solid black lines are the fits adopted in the analytic calculation with the GSM. The cyan line corresponding to the right y-axis of the top panel refers to the total proper pairwise velocity, $v_{\rm hubble} + \langle v_{12}(r_{12})\rangle$, and the black dashed line is the Hubble flow velocity $v_{\rm hubble} = H(z_s)r/(1+z_s)$. This figure shows the presence of cosmological inflow of absorbers on to galaxies, which slows down the total outflowing velocity including the Hubble flow.

$\langle v_{12}(r_{12})\rangle = -v_{in}/[1 + (r_{12}/r_{\rm v})^{\gamma_{\rm v}}]$ where $v_{in} = 133$ km s^{-1}, $r_{\rm v} = 6.3\,h^{-1}$ cMpc and $\gamma_{\rm v} = 6.2$. This is adopted to evaluate the GSM in Fig. 10. For simplicity, rather than using a fit to the curve, we assume a constant pairwise velocity dispersion equal to its mean, i.e. $\sigma_{12} = 100$ km s^{-1}.

As shown in Fig. 10, the impact of the galaxy–absorber correlation function in velocity space provides an additional boost of effective optical depth relative to the model with $\xi_{\rm v} = 0$. In fact, the enhanced clustering of absorbers around galaxies (Fig. 13) renders the IGM more opaque. Furthermore, the cosmological inflow of absorbers on to galaxies (Fig. 14) causes a departure from the Hubble flow in the immediate surroundings of galaxies and enhances the velocity–space clustering (the slower the total outflow velocity in the proper unit is, the more opaque to Lyα photons the gas becomes, as it is less redshifted out of resonance). This can increase $\tau^{\rm eff}_{\rm web}$, preferentially at the lower Δv. Thus, the effective optical depth including a velocity–space galaxy–absorber clustering is larger and steeper than the one including only a change in the CDDF amplitude (with $\xi_{\rm v} = 0$).

Overall, Fig. 10 shows that the simulation and the analytic model agree at $\Delta v > 400$ km s^{-1}, while our analytic approximation overestimates the opacity at $\Delta v < 300$ km s^{-1}, probably because we assume that the same galaxy–absorber correlation function applies to the full column density range of absorbers. This may lead to low column density absorbers with a $\xi(r_{12})$ which is too large. To address this issue, it is necessary to investigate in more detail the column density dependent clustering, the pairwise velocity field with outflow, and/or the effect of photoionization from the central galaxy.

5.4 Lyα red damping wing in web–bubble models

The top panels of Fig. 15 show the effective optical depth in the hybrid web–bubble models directly calculated from the simulations, together with the 1σ dispersion of optical depth among different lines-of-sight. Not surprisingly, the red damping wing becomes more opaque towards higher neutral fractions, and the scatter from sightline to sightline is large. The red and blue lines show the contributions to the total simulated optical depth from the bubble and web models used to build each web–bubble model, while the black lines indicate the optical depth that we obtain by simply adding the two contributions, i.e. $\tau^{\rm eff}_{\rm bub} + \tau^{\rm eff}_{\rm web}$. A comparison between the optical depth extracted from the web–bubble simulations and a sum of the optical depths extracted from the corresponding bubble and web models indicates that the two agree very well.[20]

The bottom panels of Fig. 15 show the relative contribution of small-scale absorbers and large-scale H I patches to the total damping wing opacity. We find that:

(i) In all our web–bubble models, neither component dominates the total optical depth, as $\tau^{\rm eff}_{\rm web}/\tau^{\rm eff}_{\rm bub} \sim 0.5 - 1.5$.

(ii) The relative contribution depends on the intrinsic Lyα line shift. The small-scale absorbers' contribution increases with decreasing Δv because their opacity is enhanced by the inflow on to galaxies (see Section 5.3.3). On the other hand, the H I patches are typically located at a distance $\sim$5–10 h^{-1} cMpc from Lyα-emitting galaxies (see Fig. 9), where the Hubble flow already dominates the total velocity. Therefore, they are more prominent at larger Δv.

The above two points underline the importance of correctly modelling small-scale absorbers within the large-scale bubble morphology. This section concludes our discussion on the average impact of

[20] A slight discrepancy arises because the simple sum counts twice the neutral gas outside ionized bubbles (in the form of H I patches in bubble models and small-scale absorbers in web models), while in the simulations small-scale absorbers are present only within ionized bubbles by construction. The simple sum is thus expected to result in a slightly higher optical depth.

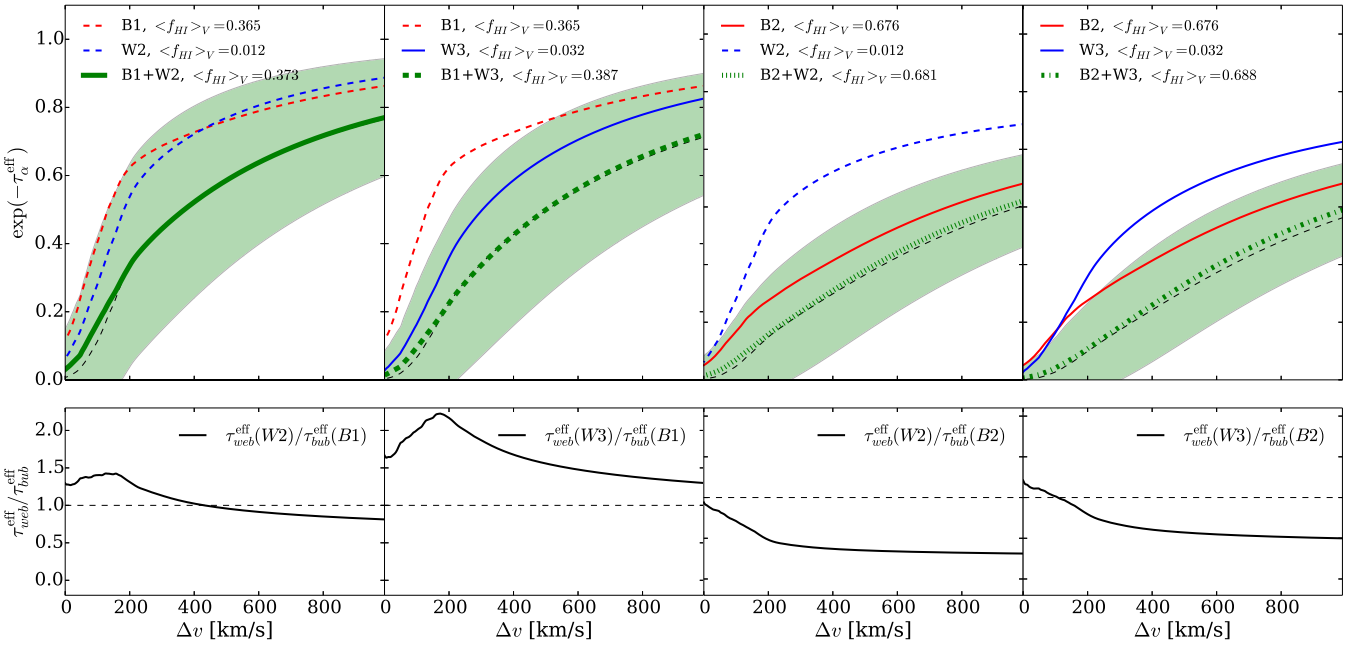

Figure 15. Top panels: Lyα effective optical depth extracted from the hybrid web–bubble model simulations (green lines; B1+W2, B1+W3, B2+W2, and B2+W3 from left to right), together with the corresponding 1σ dispersion (shaded regions). The red and blue lines are the optical depth extracted from the bubble and web models used to construct the web–bubble models, and the black lines show the sum of these two contributions, i.e. $\tau_{\rm bub}^{\rm eff} + \tau_{\rm web}^{\rm eff}$. Bottom panels: ratio of effective optical depth between web and bubble models used for the corresponding web–bubble models. This shows the impact of large-scale bubbles and small-scale absorbers on the total optical depth as a function of Δv.

large-scale neutral patches (Section 5.2) and small-scale absorbers (Section 5.3) on the Lyα red damping wing opacity in a unified web–bubble framework.

5.5 Probability distribution functions for $\mathcal{T}_{\rm IGM}$

Fig. 16 shows the distribution of the IGM transmission factor, $\mathcal{T}_{\rm IGM}$, along the line-of-sight to a Lyα-emitting galaxy as a function of the host halo mass for models B2, W3, and B1+W2. These models have been chosen because they have a similar LF (see Fig. 6) and effective optical depth (see Figs 7 and 15), and therefore a similar average Lyα visibility. The black lines are the average IGM transmission factor $\langle \mathcal{T}_{\rm IGM}(M_{\rm h})\rangle$ for each halo mass bin.

In the bubble model plotted in the top panel of Fig. 16, $\langle \mathcal{T}_{\rm IGM}(M_{\rm h})\rangle$ increases with $M_{\rm h}$, as massive [small] haloes typically reside in large [small] ionized bubbles (in the highest mass bins the trend is reversed because of the poor statistics). At the same time, there exists a population of lower mass haloes clustered around the more massive ones, which is therefore also embedded within large ionized bubbles. This explains the large scatter exhibit by $\mathcal{T}_{\rm IGM}$ for low halo masses. Furthermore, in bubble models sightlines to most (if not all) galaxies pass through H I patches, meaning that the intrinsic luminosity of most galaxies is reduced, and explaining the *unimodality* of the $\mathcal{T}_{\rm IGM}$ distribution (something that was pointed out previously by Jensen et al. 2014 and Mesinger et al. 2015).

As in web models self-shielding absorbers cluster around the more massive haloes (Section 5.3.3), $\langle \mathcal{T}_{\rm IGM}(M_{\rm h})\rangle$ decreases with increasing $M_{\rm h}$. The still present large scatter in the distribution now appears to be *bimodal*, with a peak at $\mathcal{T}_{\rm IGM} \sim 1$ and a second one at $\mathcal{T}_{\rm IGM} \sim 0$. These peaks correspond to cases in which a line-of-sight intersects an absorber or not. Differently from what happens in the bubble model where the intrinsic luminosity of all galaxies is reduced, here a suppression is [is not] present depending on whether a small-scale absorber is [is not] aligned with a galaxy, hence the bimodality. Our results are consistent with those by Mesinger et al. (2015), who also find that a bimodal distribution is a characteristic of the attenuation by small-scale absorbers.

In hybrid web–bubble models, the IGM transmission factor is a product of large-scale bubbles and small-scale absorbers. Because of the different mass-dependence of $\mathcal{T}_{\rm IGM}$ in the two models, the total IGM transmission factor here depends only weakly on $M_{\rm h}$, and no clear unimodality or bimodality in the distribution is visible. For example, the sightlines present in the web model with $\mathcal{T}_{\rm IGM} \sim 1$ are now more opaque due to the absorption from the H I patches between large-scale bubbles.

It is therefore clear that the conditional $\mathcal{T}_{\rm IGM}$-PDF at a given halo mass, $P(\mathcal{T}_{\rm IGM}|M_{\rm h})$, or in short the $\mathcal{T}_{\rm IGM}$–$M_{\rm h}$ relation, differs for web, bubble and web–bubble models. In the next section, we search for observational signatures of this variation in the intergalactic environment.

5.6 Simultaneously constraining the H I fraction and the topology of reionization

We now examine the prospect of observationally constraining the global H I fraction and the topology of reionization simultaneously by combining various statistics of Lyα-emitting galaxies.

5.6.1 The equivalent width distribution

Fig. 17 shows the cumulative probability distribution of the rest-frame equivalent width (REW), following the method of Dijkstra et al. (2011),

$$P(>\mathrm{REW}) = \int_0^1 P_{\rm intr}(>\mathrm{REW}/\mathcal{T}_{\rm IGM})\,P(\mathcal{T}_{\rm IGM})\,\mathrm{d}\mathcal{T}_{\rm IGM}, \qquad (23)$$

where the intrinsic REW distribution is $P_{\rm intr}(>\mathrm{REW}_{\rm intr}) = \exp(-\mathrm{REW}_{\rm intr}/\mathrm{REW}_{\rm c})$, with $\mathrm{REW}_{\rm c} = 50$ Å

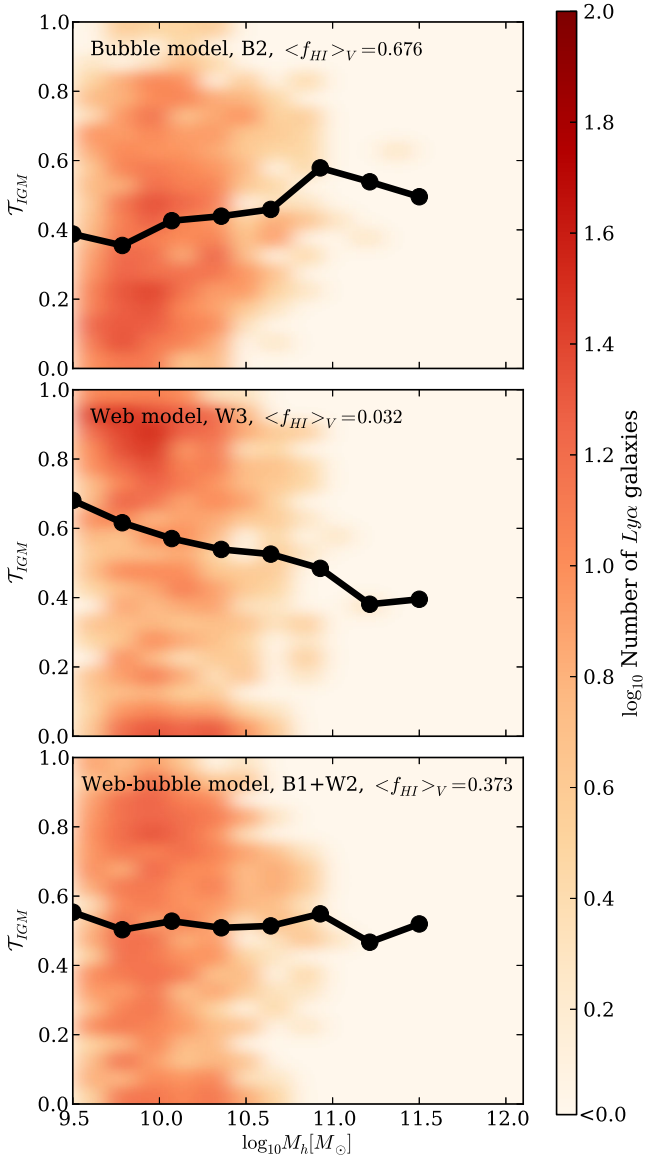

Figure 16. IGM transmission factor $\mathcal{T}_{\rm IGM}$ along the line-of-sight to a Lyα-emitting galaxy as a function of the host halo mass $M_{\rm h}$. The panels refer to the bubble model B2 (top), the web model W3 (middle), and the web–bubble model B1+W2 (bottom). The three models have a similar Lyα LF. The black lines are the average IGM transmission factors $\langle\mathcal{T}_{\rm IGM}\rangle$ in each halo mass bin. The colour indicates the number of Lyα-emitting galaxies at each location of the $M_{\rm h}-\mathcal{T}_{\rm IGM}$ map, which is a 20(log) × 20(linear) grid in the range 10^9 M$_\odot \le M_{\rm h} \le 10^{12.5}$ M$_\odot$ and $0 \le \mathcal{T}_{\rm IGM} \le 1$. The figure shows that the $\mathcal{T}_{\rm IGM}$-PDF is unimodal for a bubble model and bimodal for a web model.

(Dijkstra et al. 2011) and $\mathrm{REW_{intr}} = \mathrm{REW}/\mathcal{T}_{\rm IGM}$. The probability distribution function of the IGM transmission factor, $P(\mathcal{T}_{\rm IGM}) \propto \int P(\mathcal{T}_{\rm IGM}|M_{\rm h})\frac{{\rm d}n(>M_{\rm h})}{{\rm d}M_{\rm h}}{\rm d}M_{\rm h}$, is constructed from the simulations.

In all models, the observed REW distribution is decreased in comparison to the intrinsic one by an amount which increases with the H I fraction. Similarly to that observed for the Lyα LF, a degeneracy is present between web and bubble models, with, for example, B1 and W2 providing similar REW distributions.

However, the degeneracy can be partially broken if the REW distribution is combined with the Lyα LF. In fact, while models B2, W3 and B1+W2 are degenerate in Lyα LF (see Fig. 6) they produce distinguishable observed REW distributions. Although this is not always the case (for example, B1 and W2 show similar curves both in the Lyα LF and the REW distribution), such a combined analysis offers a test to differentiate reionization models.

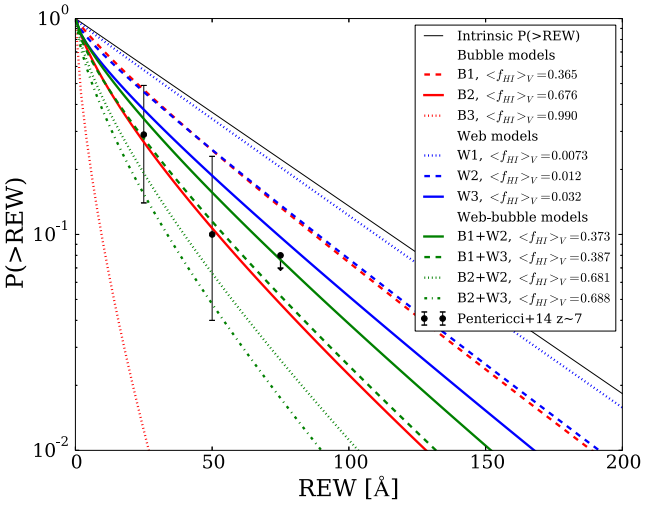

Figure 17. Cumulative probability distribution of the REW at $z = 7$. The black line is the intrinsic REW distribution and the coloured lines refer to the observed REW distributions predicted from simulations: bubble model B1 (red dashed), B2 (red solid) and B3 (red dotted); web model W1 (blue dotted), W2 (blue dashed) and W3 (blue solid); web–bubble model B1+W2 (green solid), B1+W3 (green dashed), B2+W2 (green dotted) and B2+W3 (dot–dashed). The black circles are the observation of Pentericci et al. (2014) without interloper correction (if the interloper correlation is taken into account the data points can be higher by ∼20 per cent).

The argument above can be better understood by noting that the observed Lyα LF and REW distribution depend differently on the $\mathcal{T}_{\rm IGM}$–$M_{\rm h}$ relation. To see this, we first express the Lyα LF in terms of $P(\mathcal{T}_{\rm IGM}|M_{\rm h})$ as

$$\frac{{\rm d}n(>L_\alpha^{\rm obs})}{{\rm d}L_\alpha} = \int P(L_\alpha^{\rm obs}|M_{\rm h})\frac{{\rm d}n(>M_{\rm h})}{{\rm d}M_{\rm h}}{\rm d}M_{\rm h}, \quad (24)$$

where

$$P(L_\alpha^{\rm obs}|M_{\rm h}) = \int_0^1 P_{\rm intr}(L_\alpha^{\rm obs}/\mathcal{T}_{\rm IGM}|M_{\rm h})P(\mathcal{T}_{\rm IGM}|M_{\rm h}){\rm d}\mathcal{T}_{\rm IGM}. \quad (25)$$

$P_{\rm intr}(L_\alpha|M_{\rm h})$ is the intrinsic conditional probability distribution of the Lyα luminosity given a halo mass.[21] A comparison between equations (23) and (25) shows a different dependence on $P(\mathcal{T}_{\rm IGM}|M_{\rm h})$.[22] This is because the Lyα LF is constructed from Lyα selected LAEs, while the REW-PDF is constructed from continuum selected galaxies. In fact, Dijkstra & Wyithe (2012) and Gronke et al. (2015) have shown that selection by Lyα line flux enhances the contribution of UV-faint galaxies (at fixed Lyα flux), which are absent from continuum selected samples. As such UV-faint galaxies should preferentially reside in low-mass haloes, this difference in

[21] Explicitly, we use $P_{\rm intr}(L_\alpha|M_{\rm h}) = \delta_{\rm D}[L_\alpha - L_\alpha(M_{\rm h})]$ as we assume a one-to-one mapping between L_α and $M_{\rm h}$ based on the abundance matching technique.

[22] Note that equation (23) implicitly assumes that the intrinsic REW distribution is independent of halo mass. We can, of course, generalize this modelling to include the halo mass dependence, but because this in general differs from the one of the Lyα luminosity, the dependence of the two statistics on the $\mathcal{T}_{\rm IGM}$–$M_{\rm h}$ relation is expected to differ as well.

selection function would introduce a different dependence in the $\mathcal{T}_{\rm IGM}$–$M_{\rm h}$ relation that may lead to a drop in the observed Lyα LF different from the one in the REW distribution.

Hence, a combined analysis of Lyα LF and REW distribution may allow us to constrain the H I fraction and the topology of reionization. We can already do this analysis. The upper limit at REW = 75 Å slightly favours the bubble or web–bubble models with $\langle f_{\rm H\,I}\rangle_{\rm V} \sim 68$ per cent or ∼37 per cent. If we include this constraint, the neutral fraction is favoured to be of the order of tens of per cent. This constraint is very weak because of a large uncertainty due to the interloper contamination. Moreover, the same observations favour bimodal quenching of the Lyα visibility, which is associated with web-models. This argument simply illustrates that a combined analysis of Lyα LF and REW-PDF can shed light on the history and topology of reionization.

5.6.2 *Lyα fraction of Lyman break galaxies*

The power of such joint analysis can be strengthened once the $M_{\rm UV}$-dependent Lyα fraction of LBGs measurement is included as well. The Lyα fraction of LBGs (hereafter $\mathcal{X}_{\rm Ly\alpha}$) is defined as the fraction of LBGs with a UV magnitude $M_{\rm UV}$ and Lyα REW greater than a threshold value. We generalize the method of Dijkstra et al. (2011) (see also Dijkstra & Wyithe 2012) to calculate the Lyα fraction of LBGs as

$$\mathcal{X}_{\rm Ly\alpha}(>{\rm REW}|M_{\rm UV}) = \int_0^1 \mathcal{X}^{\rm intr}_{\rm Ly\alpha}(>{\rm REW}/\mathcal{T}_{\rm IGM}|M_{\rm UV})P(\mathcal{T}_{\rm IGM}|M_{\rm UV})\,{\rm d}\mathcal{T}_{\rm IGM}, \quad (26)$$

where $\mathcal{X}^{\rm intr}_{\rm Ly\alpha}(>{\rm REW}_{\rm intr}|M_{\rm UV}) = e^{-{\rm REW}_{\rm intr}/{\rm REW}_{\rm c}(M_{\rm UV})}$ is the intrinsic Lyα fraction, ${\rm REW}_{\rm c}(M_{\rm UV})$ is a characteristic REW (see Appendix E for more details), and the conditional $\mathcal{T}_{\rm IGM}$ probability distribution function at a given $M_{\rm UV}$ is

$$P(\mathcal{T}_{\rm IGM}|M_{\rm UV}) = \int P(\mathcal{T}_{\rm IGM}|M_{\rm h})P(M_{\rm h}|M_{\rm UV})\,{\rm d}M_{\rm h}. \quad (27)$$

We construct our intrinsic model assuming that UV-bright LBGs populate more massive haloes, and consider a case with a correlation between REW and $M_{\rm UV}$ ($M_{\rm UV}$-dependent model) and one with no correlation (uncorrelated model). The $M_{\rm UV}$-dependent model is our fiducial case because observations suggest that such correlation exists (Stark et al. 2010; Jiang et al. 2013, but see Nilsson et al. 2009). More details are provided in Appendix E.

Fig. 18 shows the intrinsic and observed Lyα fractions at $z = 7$ for models B2, W3, and B1+W2, which all produce a similar Lyα LF. Two main features emerge:

(i) the bubble model shows an *upturn* of Lyα fraction at UV-bright LBGs (typically defined as those with $M_{\rm UV} < -20.25$), while the web model shows a monotonic decrease of Lyα fraction for increasing UV-bright LBGs.[23] This qualitative change in the shape of the $M_{\rm UV}$-dependent Lyα fraction is robust against different intrinsic models of REW.

(ii) In the $M_{\rm UV}$-dependent model (solid lines), the drop in the observed Lyα fraction compared to the intrinsic one is larger for UV-faint LBGs ($M_{\rm UV} > -20.25$) than for UV-bright LBGs in *all* models. The common expectation that the drop of Lyα fraction of UV-faint LBGs is larger than the one of UV-bright LBGs occurs

[23] Note that the downturn in the bubble model B2 at $M_{\rm UV} \lesssim -21.15$ is due to the poor statistics, similarly to that observed in the top panel of Fig. 16.

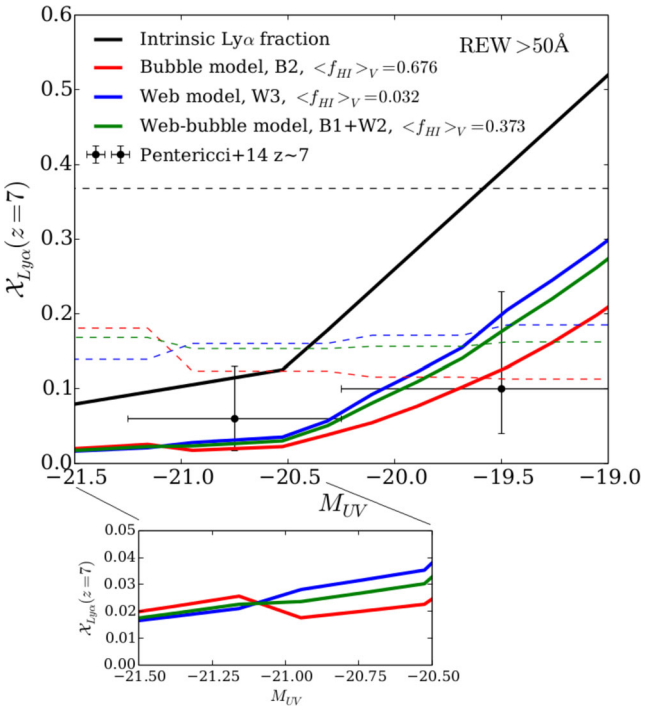

Figure 18. Lyα fraction of LBGs having REW > 50 Å at $z = 7$ as a function of the UV magnitude $M_{\rm UV}$ of galaxies. The line style indicates the intrinsic fraction obtained with the $M_{\rm UV}$-dependent (solid lines) and the uncorrelated (dashed lines) model (see text for more details), and the line colours refer to the intrinsic Lyα fraction (black), and to the observed Lyα fraction for models B2 (red), W3 (blue) and B1+W2 (green). The three models have a similar Lyα LF (Fig. 6). The black points are the observations of Pentericci et al. (2014), where the horizontal error bars indicate the bin size used. The smaller plot is a zoom-in of the Lyα fraction–$M_{\rm UV}$ relation to emphasize the upturn at UV-bright LBGs caused in model B2 by their larger Lyα visibility.

only in bubble models (Ono et al. 2012) is true *only if* the intrinsic REW and $M_{\rm UV}$ are uncorrelated (dashed lines).

The upturn of the $M_{\rm UV}$-dependent Lyα fraction can be understood as an imprint of the $\mathcal{T}_{\rm IGM}$–$M_{\rm h}$ relation (see Section 5.5). In fact, because UV-bright LBGs in bubble models are more likely to be surrounded by large ionized bubbles, the probability that their Lyα emission is visible (i.e. that they are associated with larger $\mathcal{T}_{\rm IGM}$) is higher than for UV-faint LBGs. On the other hand, in web models the small-scale absorbers cluster more strongly around UV-bright LBGs, lowering their Lyα visibility. The upturn of Lyα fraction, therefore, does not happen in web models. As a consequence, the qualitative change in the *shape* of the $M_{\rm UV}$-dependent Lyα fraction can be used as an indicator of the (possible) presence of large-scale bubbles.

On the other hand, a drop of the Lyα fraction for UV-faint LBGs larger than for UV-bright LBGs cannot be used as a decisive evidence of patchy reionization. In fact, while in the uncorrelated case (dashed lines) we indeed see a larger drop for UV-faint LBGs only for the bubble model, in the $M_{\rm UV}$-dependent case (solid lines) such drop is visible for all models. The simplest explanation for this is that because, to the first order approximation, the neutral IGM suppresses the Lyα emission by re-scaling the characteristic REW as $\langle\mathcal{T}_{\rm IGM}\rangle{\rm REW}_{\rm c}(M_{\rm UV})$ (see also Appendix E), UV-faint galaxies (with an intrinsically larger ${\rm REW}_{\rm c}$) experience a larger reduction in number above a given REW than the UV-bright galaxies (with intrinsically small ${\rm REW}_{\rm c}$) do.

In summary, the analysis of the $M_{\rm UV}$-dependent Lyα fraction of LBGs provides a powerful diagnostic tool to characterize the impact of large-scale bubbles and small-scale absorbers when properly interpreted. Hence, when combined with the Lyα LF, it offers an opportunity to constrain the H I fraction and the topology of reionization simultaneously. While the aim of the present paper is to highlight the potential of this diagnostics, we plan to use it more extensively in a future study.

6 DISCUSSION AND CONCLUSIONS

The visibility of Lyα-emitting galaxies during the EoR is controlled by both diffuse H I patches in the IGM and small-scale self-shielding absorbers around galaxies. It is therefore important to correctly include small-scale absorbers inside large-scale ionized bubbles. In this work we have explored the impact of both large-scale bubbles and small-scale absorbers on the visibility of the population of Lyα-emitting galaxies at $z > 6$, using a powerful combination of an analytic approach and hydrodynamical simulations, which covers the full range of models explored in recent investigations (Bolton & Haehnelt 2013; Jensen et al. 2013; Choudhury et al. 2015; Mesinger et al. 2015). We have considered the IGM Lyα RT in three different classes of IGM ionization structure, namely (i) the bubble model, where only large-scale ionized bubbles due to patchy reionization are present, (ii) the web model, where only small-scale absorbers are considered, and (iii) the web–bubble model, which includes both small-scale absorbers and large-scale bubbles.

Our main conclusions are as follows.

(i) The observed Lyα LF evolution from $z = 5.7$ to $z \sim 7$ requires a neutral fraction $\langle f_{\rm H\,I}\rangle_{\rm V} \sim 60$–$80$ per cent in bubble models, $\langle f_{\rm H\,I}\rangle_{\rm V} \gtrsim 1$–$3$ per cent in web models, and $\langle f_{\rm H\,I}\rangle_{\rm V} \sim 30$–$70$ per cent in web–bubble models.

(ii) A sole analysis of the Lyα LF or of the distribution of REW cannot put a stringent constraint on the reionization history. The Lyα LF and the REW-PDF can be equally suppressed in bubble, web, and web–bubble models, yet with very different global H I fractions. Hence, there is a fundamental degeneracy between the ionization structure of the IGM and the global H I fraction inferred from Lyα surveys (see Section 5.1).

(iii) We showed in Section 5.6 that a joint analysis of the Lyα LF and the REW-PDF of LBGs can improve the constraints on the neutral fraction by breaking the degeneracy with the topology of reionization.

(iv) The Lyα fraction of LBGs can be a powerful diagnostic to study the relative importance of large-scale H I patches and small-scale absorbers in the IGM. We caution that a drop in Lyα fraction that is larger for UV-faint LBGs than for UV-bright LBGs (as in Ono et al. 2012) can be reproduced with web and web–bubble models, and does not provide exclusive evidence for patchy reionization. Instead, we argue that the shape of the $M_{\rm UV}$-dependent Lyα fraction may provide more insight into the topology of reionization (see e.g. Fig. 18).

For example, an upturn of Lyα fraction for UV-bright LBGs can be caused not only by large-scale ionized bubbles, but also by an increase in the UV background around UV-bright galaxies, which reduces the abundance of small-scale absorbers. Interestingly, this upturn may already have been observed at $4.5 < z < 6$ (Stark et al. 2010), and may reflect large fluctuations in the UV background. These fluctuations have been proposed to explain observations of the cumulative effective optical depth distribution at $z \gtrsim 5$ in the spectra of high-redshift QSOs (Becker et al. 2015; Chardin et al. 2015).

(v) Our analytic formalism shows that the Lyα damping wing opacity from small-scale absorbers is highly influenced by the clustering and the pairwise velocity field of galaxy–absorber pairs (see Section 5.3). Absorbers with $N_{\rm H\,I} > 10^{19}\,{\rm cm}^{-2}$, i.e. super-LLS/DLAs, provide the largest contribution to the red damping wing at $\Delta v > 300\,{\rm km\,s}^{-1}$, while lower column density absorbers are important at smaller Δv. Understanding the galaxy–absorber correlation functions and their velocity fields can improve the robustness with which the reionization history can be constrained using Lyα-emitting galaxies. Direct observational constraints on H I CDDF and galaxy–absorbers correlation function (and as a function of $N_{\rm H\,I}$) can therefore be very useful. A possible approach is to extend to the range $3 < z < 7$ the survey strategy that searches for Lyα-emitting galaxies in the foregrounds of high-redshift QSOs, similar to the observation of Cooke et al. (2006b), Keck Baryonic Structure Survey (Rudie et al. 2012; Turner et al. 2014), and VLT LBG Redshift Survey (Crighton et al. 2011). This observational strategy is already within reach at $z \sim 5.7$ (Díaz et al. 2014).

(vi) We showed that the total effective optical depth in web–bubble models can be written as the sum of those in web and bubble models, i.e. $\tau_{\alpha}^{\rm eff} \approx \tau_{\rm bub}^{\rm eff} + \tau_{\rm web}^{\rm eff}$ (see Section 5.4). This is an important result as fast semi-numeric simulations can be used to generate $\tau_{\rm bub}^{\rm eff}$. These simulations can then be complemented with (improved) analytic or possibly empirical prescriptions for $\tau_{\rm web}^{\rm eff}$ (as in Section 5.3) to efficiently generate more realistic web–bubble models.

(vii) Web, bubble and web–bubble models produce different $\mathcal{T}_{\rm IGM}$-PDFs (Section 5.5). Bubble models show a unimodal $\mathcal{T}_{\rm IGM}$-PDF, while small-scale self-shielding absorbers in the web-model have a bimodal $\mathcal{T}_{\rm IGM}$-PDF. The modality of the hybrid web–bubble model depends on which component dominates the IGM opacity. Pentericci et al. (2014) have provided observational evidence for bimodal quenching of Lyα flux (see Treu et al. 2012, 2013 for details on the procedure). Our results imply that bimodal quenching indicates an influence of small-scale absorbers on the Lyα visibility (also see Mesinger et al. 2015), which is opposite to the common interpretation.

In conclusion, in this paper we have shown that a joint analysis of different statistics of Lyα-emitting galaxies (e.g. Lyα LF, REW distribution, Lyα fraction of LBGs, correlation function) can break degeneracies associated with individual probes. It should therefore be possible to constrain simultaneously the global H I fraction *and* the reionization topology, when armed with a suit of models of reionization in which both large-scale bubble morphology and small-scale absorbers are included.

ACKNOWLEDGEMENTS

KK thanks Hannes Jensen, Martin Haehnelt, Michele Sasdelli for useful comments and discussions, Andrew Chung for carefully reading the manuscript, Akira Konno and Masami Ouchi for kindly providing the data points shown in Fig. 6, and Romain Teyssier and the RAMSES developer team to make the code public and user friendly. We thank an anonymous referee for helpful comments.

REFERENCES

Barkana R., Loeb A., 2004, ApJ, 601, 64
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Bertschinger E., 1995, preprint, (astro-ph/9506070)
Bolton J. S., Haehnelt M. G., 2007, MNRAS, 382, 325
Bolton J. S., Haehnelt M. G., 2013, MNRAS, 429, 1695
Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS, 416, L70
Bouwens R. J. et al., 2015, ApJ, 803, 34
Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, MNRAS, 412, 2543
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, MNRAS, 443, 2831
Cassata P. et al., 2015, A&A, 573, A24
Castor J. I., 2007, in John I., ed., Radiation Hydrodynamics, Cambridge Univ. Press, Cambridge
Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, preprint, (arXiv:1505.01853)
Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, MNRAS, 452, 261
Ciardi B., Ferrara A., Marri S., Raimondo G., 2001, MNRAS, 324, 381
Ciardi B., Bolton J. S., Maselli A., Graziani L., 2012, MNRAS, 423, 558
Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006a, ApJ, 636, L9
Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006b, ApJ, 652, 994
Crighton N. H. M.et al., 2011, MNRAS, 414, 28
Curtis-Lake E. et al., 2012, MNRAS, 422, 1425
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Díaz C. G., Koyama Y., Ryan-Weber E. V., Cooke J., Ouchi M., Shimasaku K., Nakata F., 2014, MNRAS, 442, 946
Dijkstra M., 2014, PASA, 31, 40
Dijkstra M., 2016, in Mesinger A., ed., Astrophysics and Space Science Library Vol. 423. Understanding the Epoch of Cosmic Reionization: Challenges and Progress, p. 145
Dijkstra M., Wyithe J. S. B., 2010, MNRAS, 408, 352
Dijkstra M., Wyithe J. S. B., 2012, MNRAS, 419, 3181
Dijkstra M., Lidz A., Wyithe J. S. B., 2007, MNRAS, 377, 1175
Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, MNRAS, 414, 2139
Dressler A., Henry A., Martin C. L., Sawicki M., McCarthy P., Villaneuva E., 2015, ApJ, 806, 19
Eisenstein D. J., Hut P., 1998, ApJ, 498, 137
Ellis R. S. et al., 2013, ApJ, 763, L7
Faisst A. L., Capak P., Carollo C. M., Scarlata C., Scoville N., 2014, ApJ, 788, 87
Fan X. et al., 2006, AJ, 132, 117
Forero-Romero J. E., Yepes G., Gottlöber S., Prada F., 2012, MNRAS, 419, 952
Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031
Graziani L., Maselli A., Ciardi B., 2013, MNRAS, 431, 722
Gronke M., Dijkstra M., Trenti M., Wyithe S., 2015, MNRAS, 449, 1284
Haardt F., Madau P., 1996, ApJ, 461, 20
Hinshaw G. et al., 2013, ApJS, 208, 19
Hu E. M., Cowie L. L., Barger A. J., Capak P., Kakazu Y., Trouille L., 2010, ApJ, 725, 394
Hutter A., Dayal P., Partl A. M., Müller V., 2014, MNRAS, 441, 2861
Jeeson-Daniel A., Ciardi B., Maio U., Pierleoni M., Dijkstra M., Maselli A., 2012, MNRAS, 424, 2193
Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, MNRAS, 428, 1366
Jensen H., Hayes M., Iliev I. T., Laursen P., Mellema G., Zackrisson E., 2014, MNRAS, 444, 2114
Jiang L. et al., 2013, ApJ, 772, 99
Kashikawa N.et al., 2006, ApJ, 637, 631
Kim T.-S., Carswell R. F., Cristiani S., D'Odorico S., Giallongo E., 2002, MNRAS, 335, 555
Konno A. et al., 2014, ApJ, 797, 16
Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, ApJ, 728, 52
Loeb A., Furlanetto S. R., 2013, The First Galaxies in the Universe. Princeton Univ. Press, Princeton, NJ
McGreer I. D., Mesinger A., Fan X., 2011, MNRAS, 415, 3237
McLure R. J. et al., 2011, MNRAS, 418, 2074
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 381, 75
McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Dutta S., 2008, MNRAS, 388, 1101
Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648
Malhotra S., Rhoads J. E., 2004, ApJ, 617, L5
Maselli A., Ferrara A., Ciardi B., 2003, MNRAS, 345, 379
Maselli A., Ciardi B., Kanekar A., 2009, MNRAS, 393, 171
Meiksin A. A., 2009, Rev. Modern Phys., 81, 1405
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S. R., 2008, MNRAS, 385, 1348
Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566
Mihalas D., Mihalas B. W., 1984, Foundations of Radiation Hydrodynamics, Oxford Univ. Press, New York
Miralda-Escude J., 1998, ApJ, 501, 15
Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
Nilsson K. K., Möller-Nilsson O., Møller P., Fynbo J. P. U., Shapley A. E., 2009, MNRAS, 400, 232
O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137
Ono Y. et al., 2012, ApJ, 744, 83
Ouchi M. et al., 2008, ApJS, 176, 301
Ouchi M. et al., 2010, ApJ, 723, 869
Paresce F., McKee C. F., Bowyer S., 1980, ApJ, 240, 387
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Pentericci L. et al., 2014, ApJ, 793, 113
Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103
Planck Collaboration XVI et al., 2014, A&A, 571, A16
Planck Collaboration et al., 2015, preprint, (arXiv:e-prints)
Pritchard J. R., Loeb A., 2012, Rep. Progress Phys., 75, 086901
Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, 635, 123
Rahmati A., Pawlik A. H., Raicevic M., Schaye J., 2013, MNRAS, 430, 2427
Robertson B. E. et al., 2013, ApJ, 768, 71
Rudie G. C. et al., 2012, ApJ, 750, 67
Santos M. R., 2004, MNRAS, 349, 1137
Schaye J., 2001, ApJ, 559, 507
Schroeder J., Mesinger A., Haiman Z., 2013, MNRAS, 428, 3058
Seager S., Sasselov D. D., Scott D., 1999, ApJ, 523, L1
Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, MNRAS, 408, 1628
Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289
Taylor J., Lidz A., 2014, MNRAS, 437, 2542
Teyssier R., 2002, A&A, 385, 337
Tilvi V. et al., 2014, ApJ, 794, 5
Totani T., Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, PASJ, 58, 485
Treu T., Schmidt K. B., Trenti M., Bradley L. D., Stiavelli M., 2013, ApJ, 775, L29
Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, MNRAS, 445, 794
Verhamme A., Schaerer D., Maselli A., 2006, A&A, 460, 397
Willott C. J., Carilli C. L., Wagg J., Wang R., 2015, ApJ, 807, 180
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2010, ApJ, 716, 574
Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2011, ApJ, 726, 38

APPENDIX A: THE MASS-WEIGHTED NEUTRAL FRACTION IN THE POST-REIONIZED UNIVERSE

The mass-weighted H I fraction in the post-reionized universe can be estimated from DLA/LLS surveys and Lyα forest observations,

which measure the H I CDDF. As follows, this quantity can then be converted into the H I fraction embedded as Lyα absorbers, such as DLA, LLS, and diffuse IGM.

The proper number density of H I gas in the universe, $n_{\rm H_I}^{\rm prop}(z)$, is expressed as (cf. Meiksin 2009)

$$n_{\rm H_I}^{\rm prop}(z) = \int N_{\rm H_I} \frac{\partial^2 \mathcal{N}}{\partial N_{\rm H_I} \partial z} \left| \frac{{\rm d}z}{{\rm d}l_{\rm p}} \right| {\rm d}N_{\rm H_I},$$
$$= \frac{(1+z)^3 H_0}{c} \int N_{\rm H_I} f(N_{\rm H_I}, z) {\rm d}N_{\rm H_I}, \quad \text{(A1)}$$

where $l_{\rm p}$ is the proper distance, ${\rm d}l_{\rm p}/{\rm d}z = c/H(z)(1+z)$. Therefore, the fraction of neutral hydrogen over the total hydrogen atoms in the entire universe, $\langle f_{\rm H_I} \rangle_{\rm M}$, is given by $\langle f_{\rm H_I} \rangle_{\rm M} = n_{\rm H_I}^{\rm prop}(z)/\bar{n}_{\rm H}^{\rm prop}(z)$,[24]

$$\langle f_{\rm H_I} \rangle_{\rm M} = \frac{8\pi G m_{\rm H}}{3H_0 c(1-Y)\Omega_{\rm b}} \int_{N_{\rm H_I}^{\rm min}}^{N_{\rm H_I}^{\rm max}} N_{\rm H_I} f(N_{\rm H_I}, z) {\rm d}N_{\rm H_I}, \quad \text{(A2)}$$

where $m_{\rm H}$ is the mass of a hydrogen atom and $\bar{n}_{\rm H}^{\rm prop}(z) = \frac{3H_0^2(1-Y)\Omega_{\rm b}}{8\pi G m_{\rm H}}(1+z)^3 = 2.057 \times 10^{-7}(1+z)^3 (\frac{\Omega_{\rm b}h^2}{0.023}) {\rm cm}^{-3}$ for a helium abundance $Y = 0.25$. The upper and lower limits of the integration specify whether the H I content is embedded in the Lyα forest absorbers ($\log_{10}[N_{\rm H_I}/{\rm cm}^{-2}] < 17$), LLSs ($17 < \log_{10}[N_{\rm H_I}/{\rm cm}^{-2}] < 20.3$), or damped Ly$\alpha$ systems ($20.3 < \log_{10}[N_{\rm H_I}/{\rm cm}^{-2}]$). We integrate equation (A2) using the fitting functions to the observed CDDFs, $f(N_{\rm H_I}, z)$. We use the CDDF fitting functions from Kim et al. (2002) for the Lyα forest absorbers, Péroux et al. (2003) for the LLS range, and Prochaska, Herbert-Fort & Wolfe (2005) for the DLA range. The observed $f(N_{\rm H_I}, z)$ and the various fits are shown in Fig. A1.

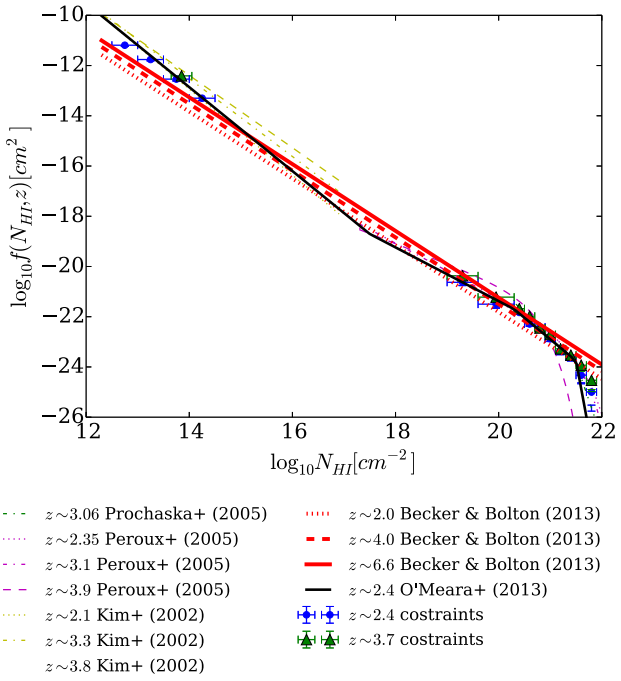

Figure A1. H I CDDF $f(N_{\rm H_I}, z)$ at $z \gtrsim 2$. The lines show the fits to the CDDF taken from the literature. The fit by Becker & Bolton (2013) is used to extrapolate to $z > 4$. The points at $z \sim 2.4$ and $z \sim 3.7$ use the compilation of data presented in O'Meara et al. (2013).

APPENDIX B: EFFECTIVE OPTICAL DEPTH OF DYNAMICAL SMALL-SCALE ABSORBERS

The opacity from small-scale absorbers is determined by the phase-space distribution function of galaxy–absorber pairs, $f(r_{12}, v_{12}, N_{\rm H_I})$, where r_{12} is the comoving separation and v_{12} is the peculiar pairwise radial velocity of pairs.

The line transfer is sensitive to the clustering in total velocity space, $v_c = aHr_{12} + v_{12}$. The probability to find an absorber within v_c and $v_c + {\rm d}v_c$ and column density $N_{\rm H_I}$ and $N_{\rm H_I} + {\rm d}N_{\rm H_I}$ is $p(v_c, N_{\rm H_I}){\rm d}v_c{\rm d}N_{\rm H_I}$. Then, the effective optical depth is given by (Paresce, McKee & Bowyer 1980)

$$\tau_{\rm web}^{\rm eff} = \iint p(v_c, N_{\rm H_I}) \left[1 - e^{-\tau_{\rm abs}(v_c, N_{\rm H_I})}\right] {\rm d}v_c {\rm d}N_{\rm H_I}. \quad \text{(B1)}$$

$p(v_c, N_{\rm H_I})$ is related to the phase-space distribution function of galaxy–absorber pairs through the transformation of variables r_{12}, v_{12} to v_c,

$$p(v_c, N_{\rm H_I}) = \iint \delta_{\rm D}\left[v_c - (aHr_{12} + v_{12})\right] f(r_{12}, v_{12}, N_{\rm H_I}) {\rm d}v_{12} {\rm d}r_{12}$$
$$= \int p_{\rm v}(v_c - aHr_{12} | r_{12}, N_{\rm H_I}) p_r(r_{12}, N_{\rm H_I})\, {\rm d}r_{12}, \quad \text{(B2)}$$

where $\delta_{\rm D}$ is the Dirac delta function. For the second equality, we have used $f(r_{12}, v_{12}, N_{\rm H_I}) = p_{\rm v}(v_{12}|r_{12}, N_{\rm H_I}) p_r(r_{12}, N_{\rm H_I})$, where $p_{\rm v}(v_{12}|r_{12}, N_{\rm H_I}){\rm d}v_{12}$ is the conditional probability to find an absorber with peculiar pairwise velocity between v_{12} and $v_{12} + {\rm d}v_{12}$ at given pair separation r_{12} and column density $N_{\rm H_I}$, and $p_r(r_{12}, N_{\rm H_I}){\rm d}r_{12}{\rm d}N_{\rm H_I}$ is the probability to find an absorber in the range r_{12} to $r_{12} + {\rm d}r_{12}$ and $N_{\rm H_I}$ to $N_{\rm H_I} + {\rm d}N_{\rm H_I}$. The real-space correlation function $\xi(r_{12}, N_{\rm H_I})$ of absorbers around galaxies gives

$$p_r(r_{12}, N_{\rm H_I}) = \frac{\partial^2 \mathcal{N}}{\partial N_{\rm H_I} \partial z} \left| \frac{{\rm d}z}{{\rm d}r} \right| \left[1 + \xi(r_{12}, N_{\rm H_I})\right], \quad \text{(B3)}$$

where $|{\rm d}r/{\rm d}z| = c/H(z_s)$. Substituting into equation (B2),

$$p_{\rm v}(v_c, N_{\rm H_I}) = \frac{\partial^2 \mathcal{N}}{\partial N_{\rm H_I} \partial z} \left| \frac{{\rm d}z}{{\rm d}r} \right| \frac{1}{aH} \left[1 + \xi_{\rm v}(v_c, N_{\rm H_I})\right], \quad \text{(B4)}$$

where we have defined the absorber–galaxy correlation function in velocity space as

$$1 + \xi_{\rm v}(v_c, N_{\rm H_I})$$
$$\equiv aH \int {\rm d}r_{12} \left[1 + \xi(r_{12}, N_{\rm H_I})\right] p_{\rm v}(v_c - aHr_{12} | r_{12}, N_{\rm H_I}). \quad \text{(B5)}$$

Thus, the effective optical depth is

$$\tau_{\rm web}^{\rm eff} = \int {\rm d}N_{\rm H_I} \frac{\partial^2 \mathcal{N}}{\partial N_{\rm H_I} \partial z} \left| \frac{{\rm d}z}{{\rm d}r} \right|$$
$$\times \int \frac{{\rm d}v_c}{aH} (1 + \xi_{\rm v}(v_c, N_{\rm H_I})) \left[1 - e^{-\tau_{\rm abs}(v_c, N_{\rm H_I})}\right]. \quad \text{(B6)}$$

All the quantities are evaluated at redshift $z = z_s$. By rearranging we obtain equation (14).

In the absence of clustering, $\xi_{\rm v} = 0$, the effective optical depth (B6) reduces to the well-known expression for the Poisson-distributed absorbers $\tau_{\rm web}^{\rm eff} = \int {\rm d}z \int {\rm d}N_{\rm H_I} |\frac{{\rm d}l_{\rm p}}{{\rm d}z}| \frac{\partial^2 \mathcal{N}}{\partial N_{\rm H_I} \partial l_{\rm p}} (1 - e^{-\tau_{\rm abs}})$ (e.g. Haardt & Madau 1996).

[24] The fraction of total number of neutral hydrogen, $\mathcal{N}_{\rm H_I}$, over the total hydrogen atom counts, $\mathcal{N}_{\rm H}$, is given by the mass-weighted neutral fraction $\langle f_{\rm H_I} \rangle_{\rm M} = \mathcal{N}_{\rm H_I}/\mathcal{N}_{\rm H} = \int x_{\rm H_I} n_{\rm H} {\rm d}V / \int n_{\rm H} {\rm d}V = \int x_{\rm H_I} \rho {\rm d}V / \int \rho {\rm d}V$. The volume-weighted and the mass-weighted neutral fraction are identical only for a homogeneous IGM: $\langle f_{\rm H_I} \rangle_{\rm M} = \int x_{\rm H_I} \bar{\rho} {\rm d}V / \int \bar{\rho} {\rm d}V = \int x_{\rm H_I} {\rm d}V / \int {\rm d}V = \langle f_{\rm H_I} \rangle_{\rm V}$.

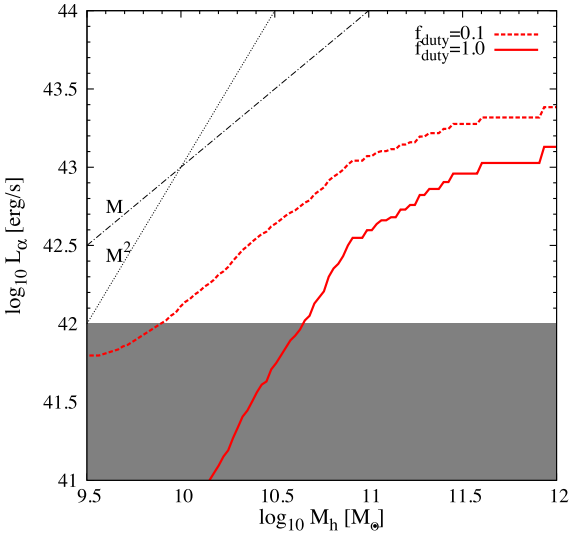

Figure C1. Semi-empirical relation between the halo mass and the intrinsic Lyα luminosity from the abundance matching technique. The red solid (dashed) line is the result of matching the simulated halo mass function at $z = 7$ with the $z = 5.7$ observed Lyα LF for $f_{\rm duty} = 1.0(0.1)$. The shadowed region indicates the luminosity range below detection limit. The two black dashed lines correspond to $L_\alpha \propto M_{\rm h}, M_{\rm h}^2$.

We show two examples of the velocity–space correlation function $\xi_{\rm v}$. For a pure Hubble flow $v_c = aHr_{12}$, $p_{\rm v}(v_{12}|r_{12}, N_{\rm H\,I}) = \delta_{\rm D}(v_{12})$. Thus, $\xi_{\rm v}(v_c) = \xi(r_{12} = \frac{v_c}{aH})$. Furthermore, a GSM is a simple generalization where the conditional pairwise peculiar velocity PDF is modelled as $p_{\rm v}(v_{12}|r_{12}, N_{\rm H\,I}) = \frac{1}{\sqrt{2\pi\sigma_{12}^2(r_{12})}} \exp[-\frac{(v_{12}-\langle v_{12}(r_{12})\rangle)^2}{2\sigma_{12}^2(r_{12})}]$, where $\langle v_{12}(r_{12})\rangle$ and $\sigma_{12}(r_{12})$ are the radial pairwise mean peculiar velocity and velocity dispersion, respectively.

APPENDIX C: ABUNDANCE MATCHING

The abundance matching technique gives a semi-empirical relation between the halo mass and the Lyα luminosity for each $f_{\rm duty}$ as shown in Fig. C1. The red lines are the result of matching the simulated halo mass function at $z = 7$ with the observed $z = 5.7$ Lyα LF (Ouchi et al. 2008) assuming a duty cycle $f_{\rm duty} = 0.1$ and 1.

Fig. C1 shows that, given a halo mass, a higher duty cycle requires a brighter Lyα luminosity to match the observed $z = 5.7$ Lyα LF, and that a simple functional form, e.g. $L_\alpha \propto M_{\rm h}, M_{\rm h}^2$, cannot match the semi-empirical relation.

In our model, the intrinsic Lyα luminosity of each galaxy (halo) is assigned according to the L_α–$M_{\rm h}$ relation with $f_{\rm duty} = 1$ in Fig. C1.

APPENDIX D: LYα RT THROUGH THE IGM: COMPUTING THE LINE-OF-SIGHT SKEWERS AND OPTICAL DEPTH

We compute the Lyα optical depth in the red damping wing as follows. The density, temperature, velocity and local H I fraction fields along skewers originating at the location of haloes and parallel to the z-axis are extracted from the hydrodynamical and radiative transfer simulations. To obtain a converged numerical integration of the optical depth, the sampling size of the skewers, δl, must be sufficiently fine. To be on the safe side, the Doppler core of the Voigt line profile should be resolved. In the velocity space this is $\delta v/c = \Delta\nu_{\rm D}/\nu_\alpha = 4.286 \times 10^{-7}(T/{\rm K})^{1/2}$. Therefore, the velocity space resolution must be $\delta v \approx 0.13(T/1{\rm K})^{1/2}\,{\rm km\,s^{-1}}$, which corresponds to a real space resolution of $\delta l = \delta v/H(z_s) \approx 0.17(T/{\rm K})^{1/2}$~pkpc at $z_s = 7$ with our cosmological parameters. If this criterion is not met, scattering by Doppler core could be missed. Although the Doppler core scattering is important in low-density regions to produce Lyα forest absorption blueward of the rest-frame Lyα line, here we are interested only in the red damping wing and the Lorentz wing scattering. Therefore, a converged evaluation of the optical depth in the red damping wing can still be obtained without strictly meeting this resolution criterion. Nonetheless, the sampling of the line-of-sight skewers must be sufficiently fine, and a sub-sampling within a cell of the cosmological hydrodynamical simulations is required to obtain a convergence in equation (D1).

To this aim, we have assumed that the density, ionization, temperature and peculiar velocity fields are constant within each cell, while the Hubble flow is allowed to vary. This is required to recover the analytic solution and to obtain a numerically converged optical depth in the limit of homogeneous expanding IGM.

The discretized form of the optical depth is then integrated at each frequency point $\nu_{\rm e}$ using the line-of-sight skewers according to

$$\tau_\alpha(\nu_{\rm e}) = \sum_{i=1}^{N} \sigma_\alpha n_{\rm H\,I}(l_i)\varphi_\nu \left[T_i, \nu_{\rm e}\left(1 - \frac{v_{\rm tot}(l_i)}{c}\right)\right] \delta l. \quad \text{(D1)}$$

The maximum proper length of the line-of-sight skewers influences the far redward optical depth, as a lower length would result in more transmission. We choose the maximum proper length of the skewer to be 12 pMpc. If a skewer exits the simulation box, a random cell in a random face of the box is chosen, and the line-of-sight is followed until the maximum proper length is reached. We have verified that for a homogeneous expanding IGM, the result at $\Delta v \sim 1000\,{\rm km\,s^{-1}}$ has a discrepancy of ~8 per cent relative to the analytic solution of the optical depth. Because the IGM will become more ionized as Lyα photons travel through the medium and because we retain the same redshift output to extract the line-of-sight skewers, we choose the maximum length of our skewer samples to be 12 pMpc.

The lower bound of the optical depth integration is chosen to be $300\,h^{-1}$ ckpc. As a reference, the virial radius of a halo with mass $M_{\rm h}$ is $R_{\rm vir} \approx 78.5(M_{\rm h}/10^{11}\,h^{-1}{\rm M}_\odot)^{1/3}\,h^{-1}$ ckpc, i.e. we exclude from the calculation the gas contained within a halo, as well as all the structures on scales smaller than the Jeans length because they are not well resolved in our simulations.

APPENDIX E: INTRINSIC Lyα FRACTION

We write the intrinsic Lyα fraction as $\mathcal{X}_{\rm Ly\alpha}^{\rm intr}(>{\rm REW_{intr}}|M_{\rm UV}) = e^{-{\rm REW_{intr}}/{\rm REW_c}(M_{\rm UV})}$ where ${\rm REW_c}(M_{\rm UV})$ is the characteristic REW.

The $M_{\rm UV}$-dependent model and uncorrelated model differ in their functional form of ${\rm REW_c}(M_{\rm UV})$, as the latter assumes a constant ${\rm REW_c}(M_{\rm UV}) = 50$ Å, while the former uses the ${\rm REW_c}(M_{\rm UV})$ obtained from the best fit to the Lyα fraction of LBGs observed at $3 < z < 6$ (Stark et al. 2010), i.e. $\mathcal{X}_{\rm Ly\alpha}^{\rm intr}(>{\rm REW}|M_{\rm UV}, z = 7) = \mathcal{X}_{\rm Ly\alpha}^{3<z<6}(>{\rm REW}|M_{\rm UV})$.

Furthermore, for $P(M_{\rm h}|M_{\rm UV})$ we assume a one-to-one mapping between UV magnitude and halo mass, i.e. $P(M_{\rm h}|M_{\rm UV}) = \delta_{\rm D}(M_{\rm h} - M_{\rm h}(M_{\rm UV}))$. The $M_{\rm h}$–$M_{\rm UV}$ relation is given by $M_{\rm h}(M_{\rm UV}) = M_{\rm h}^* \times 10^{-(M_{\rm UV}-M_{\rm UV}^*)/2.5}$ where $M_{\rm h}^* = 10^{10}\,{\rm M}_\odot$ and $M_{\rm UV}^* = -19$. The assumed scaling between $M_{\rm UV}$ and $M_{\rm h}$ is compared to the semi-empirical relation from the abundance matching with the UV LF (Bouwens et al. 2015). We find that in the range

$-21 < M_{\rm UV} < -19$ the linear scaling describes well the semi-empirical relation. However, we note that this relation tends to assign masses which are typically lower than those derived from observations. For example, $M_{\rm h}(M_{\rm UV} = -20) = 2.5 \times 10^{10}\,{\rm M}_{\odot}$, which is much lower than the mass of LBGs hosts inferred from clustering analysis, i.e. $M_{\rm h} \sim 3 \times 10^{11} - 10^{12}\,{\rm M}_{\odot}$ (e.g. Kashikawa et al. 2006). Since we expect the dependence of $\mathcal{T}_{\rm IGM}$ on halo mass to extend in the range $11 < \log_{10} M_{\rm h}/{\rm M}_{\odot} < 12$, we assume the sampling of the $\mathcal{T}_{\rm IGM}$–$M_{\rm h}$ relation at low-mass haloes to mimic the realistic host halo mass of observed LBGs.

This paper has been typeset from a TEX/LATEX file prepared by the author.